# Submodular Functions, Optimization, and Applications to Machine Learning

## — Spring Quarter, Lecture 2 —

http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/

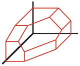

Prof. Jeff Bilmes

University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes


April 3rd, 2014

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$

$= f(A_r) + 2f(C) + f(B_r)$ $\quad = f(A_r) + f(C) + f(B_r)$ $\quad = f(A \cap B)$



# Submodular Functions, Optimization, and Applications to Machine Learning

## — Spring Quarter, Lecture 2 —

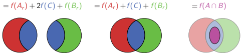
http://j.ee.washington.edu/~bilmes/classes/ee596b_spring_2014/

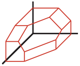

Prof. Jeff Bilmes

University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes


April 3rd, 2014

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$



# Cumulative Outstanding Reading

- Read chapter 1 from Fujishige book.

# Announcements, Assignments, and Reminders

- our room (Mueller Hall Room 154) is changed!
- Please do use our discussion board (`https://canvas.uw.edu/courses/895956/discussion_topics`) for all questions, comments, so that all will benefit from them being answered.
- Weekly Office Hours: Wednesdays, 5:00-5:50, or by skype or google hangout (email me).

# Class Road Map - IT-I

- L1 (3/31): Motivation, Applications, & Basic Definitions
- L2: (4/2): Applications, Basic Definitions, Properties
- L3:
- L4:
- L5:
- L6:
- L7:
- L8:
- L9:
- L10:
- L11:
- L12:
- L13:
- L14:
- L15:
- L16:
- L17:
- L18:
- L19:
- L20:

Finals Week: June 9th-13th, 2014.

# Submodular Definitions

**Definition 2.2.2 (submodular concave)**

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \quad (2.2)$$

An alternate and (as we will soon see) equivalent definition is:

**Definition 2.2.3 (diminishing returns)**

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A \subseteq B \subset V$, and $v \in V \setminus B$, we have that:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \quad (2.3)$$

This means that the incremental "value", "gain", or "cost" of $v$ decreases (diminishes) as the context in which $v$ is considered grows from $A$ to $B$.

# Example Discrete Optimization Problems

- **Combinatorial Problems**: e.g., set cover, max $k$ coverage, vertex cover, edge cover, graph cut problems.
- **Operations Research**: facility location (uncapacited)
- **Sensor placement**
- **Information:** Information gain and feature selection, information theory
- **Mathematics:** e.g., monge matrices
- **Networks**: Social networks, influence, viral marketing, information cascades, diffusion networks
- **Graphical models**: tree distributions, factors, and image segmentation
- **Diversity** and its models
- **NLP**: Natural language processing: document summarization, web search, information retrieval
- **ML**: Machine Learning: active/semi-supervised learning
- **Economics**: markets, economies of scale

# Markets: Supply Side Economies of scale

- Economies of Scale: Many goods and services can be produced at a much lower per-unit cost only if they are produced in very large quantities.
- The profit margin for producing a unit of goods is improved as more of those goods are created.
- If you already make a good, making a similar good is easier than if you start from scratch (e.g., Apple making both iPod and iPhone).
- An argument in favor of free trade is that it opens up larger markets for firms (especially in otherwise small markets), thereby enabling better economies of scale, and hence greater efficiency (lower costs and resources per unit of good produced).

# Supply Side Economies of scale

- What is a good model of the cost of manufacturing a set of items?

# Supply Side Economies of scale

- What is a good model of the cost of manufacturing a set of items?
- Let $V$ be a set of possible items that a company might possibly wish to manufacture, and let $f(S)$ for $S \subseteq V$ be the cost to that company to manufacture subset $S$.

# Supply Side Economies of scale

- What is a good model of the cost of manufacturing a set of items?
- Let $V$ be a set of possible items that a company might possibly wish to manufacture, and let $f(S)$ for $S \subseteq V$ be the cost to that company to manufacture subset $S$.
- Ex: $V$ might be colors of paint in a paint manufacturer: green, red, blue, yellow, white, etc.

# Supply Side Economies of scale

- What is a good model of the cost of manufacturing a set of items?
- Let $V$ be a set of possible items that a company might possibly wish to manufacture, and let $f(S)$ for $S \subseteq V$ be the cost to that company to manufacture subset $S$.
- Ex: $V$ might be colors of paint in a paint manufacturer: green, red, blue, yellow, white, etc.
- Producing green when you are already producing yellow and blue is probably cheaper than if you were only producing some other colors.

$$f(\text{green}, \text{blue}, \text{yellow}) - f(\text{blue}, \text{yellow}) <= f(\text{green}, \text{blue}) - f(\text{blue}) \tag{2.1}$$

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.
- the value of a network to a user depends on the number of other users in that network. External use benefits internal use.

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.
- the value of a network to a user depends on the number of other users in that network. External use benefits internal use.
- This is called network externalities (Katz & Shapiro 1986), and is a form of "demand" economies of scale

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.
- the value of a network to a user depends on the number of other users in that network. External use benefits internal use.
- This is called network externalities (Katz & Shapiro 1986), and is a form of "demand" economies of scale
- "value" in this case can be seen as a "willingness-to-pay" for the service (WTP)

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.
- the value of a network to a user depends on the number of other users in that network. External use benefits internal use.
- This is called network externalities (Katz & Shapiro 1986), and is a form of "demand" economies of scale
- "value" in this case can be seen as a "willingness-to-pay" for the service (WTP)
- WTP tends to increase but then saturate (like a logistic function)

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.
- the value of a network to a user depends on the number of other users in that network. External use benefits internal use.
- This is called network externalities (Katz & Shapiro 1986), and is a form of "demand" economies of scale
- "value" in this case can be seen as a "willingness-to-pay" for the service (WTP)
- WTP tends to increase but then saturate (like a logistic function)
- Given network externalities, a consumer in today's market cares also about the future success of the product and competing products.

# Demand side Economies of Scale: Network Externalities

- consumers of a good derive positive value when size of the market increases.
- the value of a network to a user depends on the number of other users in that network. External use benefits internal use.
- This is called network externalities (Katz & Shapiro 1986), and is a form of "demand" economies of scale
- "value" in this case can be seen as a "willingness-to-pay" for the service (WTP)
- WTP tends to increase but then saturate (like a logistic function)
- Given network externalities, a consumer in today's market cares also about the future success of the product and competing products.
- If the good is durable (e.g., a car or phone) or there is human capital investment (e.g., education in a skill), the total benefits derived from a good will depend on the number of consumers who adopt compatible products in the future.

# Positive Network Externalities

- railroad - standard rail format and shared access

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online
- online education, Massive Open Online Courses (MOOCs) such as Coursera, edX, etc. – with many people simultaneously taking a class, all gain due to richer peer discussions due to greater pool of well matched study groups, more simultaneous similar questions/problems that are asked $\Rightarrow$ more efficient learning & training data for ML algorithms to learn how people learn.

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online
- online education, Massive Open Online Courses (MOOCs) such as Coursera, edX, etc. – with many people simultaneously taking a class, all gain due to richer peer discussions due to greater pool of well matched study groups, more simultaneous similar questions/problems that are asked $\Rightarrow$ more efficient learning & training data for ML algorithms to learn how people learn.
- Software (e.g., Microsoft office, smartphone apps, etc.): more people means more bug reporting $\Rightarrow$ better & faster software evolution.

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online
- online education, Massive Open Online Courses (MOOCs) such as Coursera, edX, etc. – with many people simultaneously taking a class, all gain due to richer peer discussions due to greater pool of well matched study groups, more simultaneous similar questions/problems that are asked $\Rightarrow$ more efficient learning & training data for ML algorithms to learn how people learn.
- Software (e.g., Microsoft office, smartphone apps, etc.): more people means more bug reporting $\Rightarrow$ better & faster software evolution.
- gmail and web-based email (collaborative spam filtering).

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online
- online education, Massive Open Online Courses (MOOCs) such as Coursera, edX, etc. – with many people simultaneously taking a class, all gain due to richer peer discussions due to greater pool of well matched study groups, more simultaneous similar questions/problems that are asked $\Rightarrow$ more efficient learning & training data for ML algorithms to learn how people learn.
- Software (e.g., Microsoft office, smartphone apps, etc.): more people means more bug reporting $\Rightarrow$ better & faster software evolution.
- gmail and web-based email (collaborative spam filtering).
- wikipedia, collaborative documents

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online
- online education, Massive Open Online Courses (MOOCs) such as Coursera, edX, etc. – with many people simultaneously taking a class, all gain due to richer peer discussions due to greater pool of well matched study groups, more simultaneous similar questions/problems that are asked $\Rightarrow$ more efficient learning & training data for ML algorithms to learn how people learn.
- Software (e.g., Microsoft office, smartphone apps, etc.): more people means more bug reporting $\Rightarrow$ better & faster software evolution.
- gmail and web-based email (collaborative spam filtering).
- wikipedia, collaborative documents
- any widely used standard (job training now is useful in the future)

# Positive Network Externalities

- railroad - standard rail format and shared access
- The telephone, who wants to talk by phone only to oneself?
- the internet, more valuable per person the more people use it.
- ebooks (the more people comment, the better it gets)
- social network sites: facebook more valuable with everyone online
- online education, Massive Open Online Courses (MOOCs) such as Coursera, edX, etc. – with many people simultaneously taking a class, all gain due to richer peer discussions due to greater pool of well matched study groups, more simultaneous similar questions/problems that are asked $\Rightarrow$ more efficient learning & training data for ML algorithms to learn how people learn.
- Software (e.g., Microsoft office, smartphone apps, etc.): more people means more bug reporting $\Rightarrow$ better & faster software evolution.
- gmail and web-based email (collaborative spam filtering).
- wikipedia, collaborative documents
- any widely used standard (job training now is useful in the future)
- the "tipping point", and "winner take all" (one platform prevails)

# Other Network Externalities

No Network Externalities

- food/drink - (should be) independent of how many others are eating the type of food.

# Other Network Externalities

No Network Externalities

- food/drink - (should be) independent of how many others are eating the type of food.
- Music - your enjoyment should (ideally) be independent of others' enjoyment (but maybe not, see Salganik, Dodds, Watts'06).

# Other Network Externalities

No Network Externalities

- food/drink - (should be) independent of how many others are eating the type of food.
- Music - your enjoyment should (ideally) be independent of others' enjoyment (but maybe not, see Salganik, Dodds, Watts'06).

Negative Network Externalities

# Other Network Externalities

No Network Externalities

- food/drink - (should be) independent of how many others are eating the type of food.
- Music - your enjoyment should (ideally) be independent of others' enjoyment (but maybe not, see Salganik, Dodds, Watts'06).

Negative Network Externalities

- clothing

# Other Network Externalities

No Network Externalities

- food/drink - (should be) independent of how many others are eating the type of food.
- Music - your enjoyment should (ideally) be independent of others' enjoyment (but maybe not, see Salganik, Dodds, Watts'06).

Negative Network Externalities

- clothing
- (Halloween) costumes

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.
- We choose initial price $p$ and initial set of users $A \subseteq V$ who get the good for free.

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.
- We choose initial price $p$ and initial set of users $A \subseteq V$ who get the good for free.
- Define $S_1 = \{i \notin A : v_i(A) \geq p\}$ initial set of buyers.

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.
- We choose initial price $p$ and initial set of users $A \subseteq V$ who get the good for free.
- Define $S_1 = \{i \notin A : v_i(A) \geq p\}$ initial set of buyers.
- $S_2 = \{i \notin A \cup S_1 : v_i(A \cup S_1) \geq p\}$.

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.
- We choose initial price $p$ and initial set of users $A \subseteq V$ who get the good for free.
- Define $S_1 = \{i \notin A : v_i(A) \geq p\}$ initial set of buyers.
- $S_2 = \{i \notin A \cup S_1 : v_i(A \cup S_1) \geq p\}$.
- This starts a cascade. Let $S_k = \{i \notin \cup_{j<k} S_j \cup A : v_j(\cup_{j<k} S_j \cup A) \geq p\}$,

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.
- We choose initial price $p$ and initial set of users $A \subseteq V$ who get the good for free.
- Define $S_1 = \{i \notin A : v_i(A) \geq p\}$ initial set of buyers.
- $S_2 = \{i \notin A \cup S_1 : v_i(A \cup S_1) \geq p\}$.
- This starts a cascade. Let $S_k = \{i \notin \cup_{j<k} S_j \cup A : v_j(\cup_{j<k} S_j \cup A) \geq p\}$,
- and let $S_{k*}$ be the saturation point, lowest value of $k$ such that $S_k = S_{k+1}$

# Optimization Problem Involving Network Externalities

- (From Mirrokni, Roch, Sundararajan 2012): Let $V$ be a set of users.
- Let $v_i(S)$ be the value that user $i$ has for a good if $S \subseteq V$ already own the good — e.g. $v_i(S) = \omega_i + f_i(\sum_{j \in S} w_{ij})$ where $\omega_i$ is inherent value, and $f_i$ might be a concave function, and $w_{ij}$ is how important $j \in S$ is to $i$ (e.g., a network). Weights might be random.
- Given price $p$ for good, user $i$ buys good if $v_i(S) \geq p$.
- We choose initial price $p$ and initial set of users $A \subseteq V$ who get the good for free.
- Define $S_1 = \{i \notin A : v_i(A) \geq p\}$ initial set of buyers.
- $S_2 = \{i \notin A \cup S_1 : v_i(A \cup S_1) \geq p\}$.
- This starts a cascade. Let $S_k = \{i \notin \cup_{j<k} S_j \cup A : v_j(\cup_{j<k} S_j \cup A) \geq p\}$,
- and let $S_{k*}$ be the saturation point, lowest value of $k$ such that $S_k = S_{k+1}$a
- Goal: find $A$ and $p$ to maximize $f_p(A) = \mathbb{E}[p \times |S_{k^*}|]$.

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.
- Ex: Let $V = \{v_1, v_2\}$ be items with $v_1$ being the action "buy milk at the store" and $v_2$ being the action "buy honey at the store.".

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.
- Ex: Let $V = \{v_1, v_2\}$ be items with $v_1$ being the action "buy milk at the store" and $v_2$ being the action "buy honey at the store.".
- For $A \subseteq V$, let $f(A)$ be the cost of set of items $A$.

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.
- Ex: Let $V = \{v_1, v_2\}$ be items with $v_1$ being the action "buy milk at the store" and $v_2$ being the action "buy honey at the store.".
- For $A \subseteq V$, let $f(A)$ be the cost of set of items $A$.
- $f(\{v_1\})$ = cost to drive to/from store and cost to purchase milk, say $c_d + c_m$.

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.
- Ex: Let $V = \{v_1, v_2\}$ be items with $v_1$ being the action "buy milk at the store" and $v_2$ being the action "buy honey at the store.".
- For $A \subseteq V$, let $f(A)$ be the cost of set of items $A$.
- $f(\{v_1\})$ = cost to drive to/from store and cost to purchase milk, say $c_d + c_m$.
- $f(\{v_2\})$ = cost to drive to/from store and cost to purchase honey, say $c_d + c_h$.

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.
- Ex: Let $V = \{v_1, v_2\}$ be items with $v_1$ being the action "buy milk at the store" and $v_2$ being the action "buy honey at the store.".
- For $A \subseteq V$, let $f(A)$ be the cost of set of items $A$.
- $f(\{v_1\})$ = cost to drive to/from store and cost to purchase milk, say $c_d + c_m$.
- $f(\{v_2\})$ = cost to drive to/from store and cost to purchase honey, say $c_d + c_h$.
- But $f(\{v_1, v_2\}) = c_d + c_m + c_h < 2c_d + c_m + c_h$ since $c_d$ is a shared fixed cost.

# Shared Fixed Costs

- It is often inaccurate to consider individual costs in isolation, without accounting for the various interactions that might exist between them.
- Ex: Let $V = \{v_1, v_2\}$ be items with $v_1$ being the action "buy milk at the store" and $v_2$ being the action "buy honey at the store.".
- For $A \subseteq V$, let $f(A)$ be the cost of set of items $A$.
- $f(\{v_1\})$ = cost to drive to/from store and cost to purchase milk, say $c_d + c_m$.
- $f(\{v_2\})$ = cost to drive to/from store and cost to purchase honey, say $c_d + c_h$.
- But $f(\{v_1, v_2\}) = c_d + c_m + c_h < 2c_d + c_m + c_h$ since $c_d$ is a shared fixed cost.
- Shared fixed costs are submodular.

# Anecdote

From David Brooks, NYTs column, March 28th, 2011 on "Tools for Thinking". In response to Steven Pinker (Harvard) asking a number of people "What scientific concept would improve everybody's cognitive toolkit?"

> *Emergent systems are ones in which many different elements interact. The pattern of interaction then produces a new element that is greater than the sum of the parts, which then exercises a top-down influence on the constituent elements.*

# Submodular Motivation Recap

- Given a set of objects $V = \{v_1, \ldots, v_n\}$ and a function $f : 2^V \to \mathbb{R}$ that returns a real value for any subset $S \subseteq V$.
- Suppose we are interested in finding the subset that either maximizes or minimizes the function, e.g., $\text{argmax}_{S \subseteq V} f(S)$, possibly subject to some constraints.
- In general, this problem has exponential time complexity.
- Example: $f$ might correspond to the value (e.g., information gain) of a set of sensor locations in an environment, and we wish to find the best set $S \subseteq V$ of sensors locations given a fixed upper limit on the number of sensors $|S|$.
- In many cases (such as above) $f$ has properties that make its optimization tractable to either exactly or approximately compute.
- One such property is *submodularity*.

# Submodular Definitions

**Definition 2.4.2 (submodular concave)**

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) \tag{2.2}$$

An alternate and (as we will soon see) equivalent definition is:

**Definition 2.4.3 (diminishing returns)**

A function $f : 2^V \to \mathbb{R}$ is submodular if for any $A \subseteq B \subset V$, and $v \in V \setminus B$, we have that:

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B) \tag{2.3}$$

This means that the incremental "value", "gain", or "cost" of $v$ decreases (diminishes) as the context in which $v$ is considered grows from $A$ to $B$.

# Subadditive Definitions

## Definition 2.4.1 (subadditive)

A function $f : 2^V \to \mathbb{R}$ is subadditive if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \geq f(A \cup B) \tag{2.2}$$

This means that the "whole" is less than the sum of the parts.

# Supermodular Definitions

**Definition 2.4.2 (supermodular convex)**

A function $f : 2^V \to \mathbb{R}$ is supermodular if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \leq f(A \cup B) + f(A \cap B) \tag{2.3}$$

An alternate and equivalent definition is:

**Definition 2.4.3 (increasing returns)**

A function $f : 2^V \to \mathbb{R}$ is supermodular if for any $A \subseteq B \subset V$, and $v \in V \setminus B$, we have that:

$$f(A \cup \{v\}) - f(A) \leq f(B \cup \{v\}) - f(B) \tag{2.4}$$

The incremental "value", "gain", or "cost" of $v$ increases as the context in which $v$ is considered grows from $A$ to $B$.

# Submodular vs. Supermodular

- Submodular and supermodular functions are closely related.

# Submodular vs. Supermodular

- Submodular and supermodular functions are closely related.
- In fact, $f$ is submodular iff $-f$ is supermodular.

# Superadditive Definitions

**Definition 2.4.4 (superadditive)**

A function $f : 2^V \to \mathbb{R}$ is superadditive if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \leq f(A \cup B) \tag{2.5}$$

- This means that the “whole” is greater than the sum of the parts.

# Superadditive Definitions

**Definition 2.4.4 (superadditive)**

A function $f : 2^V \to \mathbb{R}$ is superadditive if for any $A, B \subseteq V$, we have that:

$$f(A) + f(B) \leq f(A \cup B) \tag{2.5}$$

- This means that the "whole" is greater than the sum of the parts.
- In general, submodular and subadditive (and supermodular and superadditive) are different properties.

# Modular Definitions

**Definition 2.4.5 (modular)**

A function that is both submodular and supermodular is called **modular**

If $f$ is a modular function, than for any $A, B \subseteq V$, we have

$$f(A) + f(B) = f(A \cap B) + f(A \cup B) \tag{2.6}$$

In modular functions, elements do not interact (or cooperate, or compete, or influence each other), and have value based only on singleton values.

**Proposition 2.4.6**

*If $f$ is modular, it may be written as*

$$f(A) = f(\emptyset) + \sum_{a \in A} \Big( f(\{a\}) - f(\emptyset) \Big) \tag{2.7}$$

# Modular Definitions

## Proof.

We inductively construct the value for $A = \{a_1, a_2, \dots, a_k\}$.

For $k = 2$,

$$f(a_1) + f(a_2) = f(a_1, a_2) + f(\emptyset) \tag{2.8}$$

$$\text{implies } f(a_1, a_2) = f(a_1) - f(\emptyset) + f(a_2) \underbrace{- f(\emptyset) + f(\emptyset)}_{0} \tag{2.9}$$

then for $k = 3$,

$$f(a_1, a_2) + f(a_3) = f(a_1, a_2, a_3) + f(\emptyset) \tag{2.10}$$

$$\text{implies } f(a_1, a_2, a_3) = f(a_1, a_2) - f(\emptyset) + f(a_3) - f(\emptyset) + f(\emptyset) \tag{2.11}$$

$$= \cancel{f(\emptyset)} + \sum_{i=1}^{3} \big(f(a_i) - \cancel{f(\emptyset)}\big) \tag{2.12}$$

and so on ... □

# Complement function

Given a function $f : 2^V \to \mathbb{R}$, we can find a complement function $\bar{f} : 2^V \to \mathbb{R}$ as $\bar{f}(A) = f(V \setminus A)$ for any $A$.

## Proposition 2.4.7

*$\bar{f}$ is submodular if $f$ is submodular.*

## Proof.

$$\bar{f}(A) + \bar{f}(B) \geq \bar{f}(A \cup B) + \bar{f}(A \cap B) \tag{2.13}$$

follows from

$$f(V \setminus A) + f(V \setminus B) \geq f(V \setminus (A \cup B)) + f(V \setminus (A \cap B)) \tag{2.14}$$

which is true because $V \setminus (A \cup B) = (V \setminus A) \cap (V \setminus B)$ and $V \setminus (A \cap B) = (V \setminus A) \cup (V \setminus B)$. □

# Submodularity

- Submodular functions have a long history in economics, game theory, combinatorial optimization, electrical networks, and operations research.
- They are gaining importance in machine learning as well (one of our main motivations for offering this course).
- Arbitrary set functions are hopelessly difficult to optimize, while the minimum of submodular functions can be found in polynomial time, and the maximum can be constant-factor approximated in low-order polynomial time.
- Submodular functions share properties in common with both convex and concave functions, but they are quite different.

# Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. *Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.*

# Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. *Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.*
2. *Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.*

# Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. *Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.*
2. *Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.*
3. *Convex functions and domains exhibit sufficient structure so that a mathematically beautiful and practically useful theory can be developed.*

# Attractions of Convex Functions

Why do we like Convex Functions? (Quoting Lovász 1983):

1. *Convex functions occur in many mathematical models in economy, engineering, and other sciences. Convexity is a very natural property of various functions and domains occurring in such models; quite often the only non-trivial property which can be stated in general.*
2. *Convexity is preserved under many natural operations and transformations, and thereby the effective range of results can be extended, elegant proof techniques can be developed as well as unforeseen applications of certain results can be given.*
3. *Convex functions and domains exhibit sufficient structure so that a mathematically beautiful and practically useful theory can be developed.*
4. *There are theoretically and practically (reasonably) efficient methods to find the minimum of a convex function.*

# Attractions of Submodular Functions

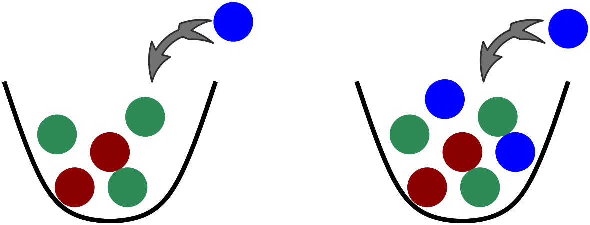
In this course, we wish to demonstrate that submodular functions also possess attractions of these four sorts as well.

# Example Submodular: Number of Colors of Balls in Urns

- Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.

# Example Submodular: Number of Colors of Balls in Urns

- Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.

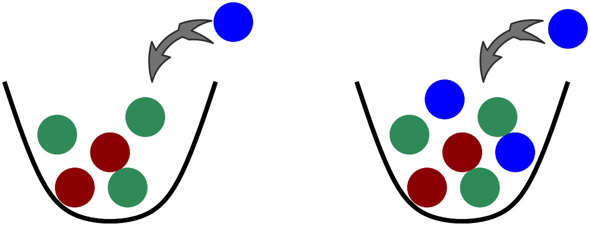


Initial value: 2 (colors in urn).
New value with added blue ball: 3

Initial value: 3 (colors in urn).
New value with added blue ball: 3

# Example Submodular: Number of Colors of Balls in Urns

- Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.



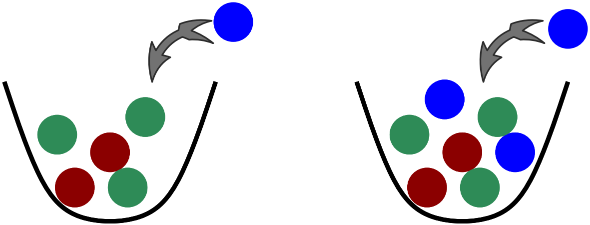
Initial value: 2 (colors in urn).
New value with added blue ball: 3

Initial value: 3 (colors in urn).
New value with added blue ball: 3

- Submodularity: Incremental Value of Object Diminishes in a Larger Context (diminishing returns).

# Example Submodular: Number of Colors of Balls in Urns

- Consider an urn containing colored balls. Given a set $S$ of balls, $f(S)$ counts the number of distinct colors.



Initial value: 2 (colors in urn).
New value with added blue ball: 3

Initial value: 3 (colors in urn).
New value with added blue ball: 3

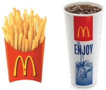
- Submodularity: Incremental Value of Object Diminishes in a Larger Context (diminishing returns).
- Thus, $f$ is submodular.

# Ex. Submodular: Consumer Costs of Living

- Consumer costs are very often submodular.

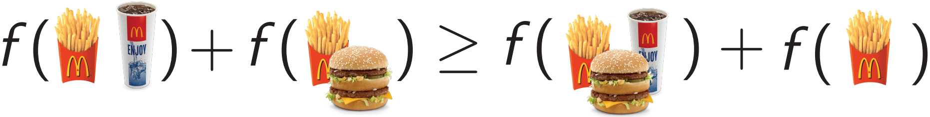
# Ex. Submodular: Consumer Costs of Living

- Consumer costs are very often submodular. For example:

$f($  $) + f($  $) \geq f($  $) + f($  $)$

# Ex. Submodular: Consumer Costs of Living

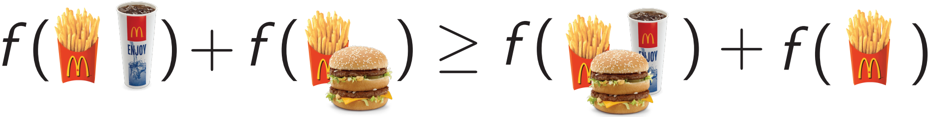
- Consumer costs are very often submodular. For example:



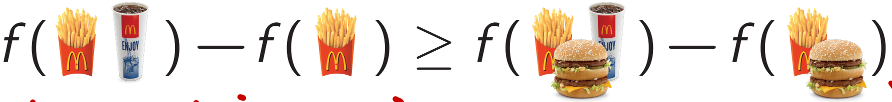
- Rearranging terms, we can see this as diminishing returns:

# Ex. Submodular: Consumer Costs of Living

- Consumer costs are very often submodular. For example:

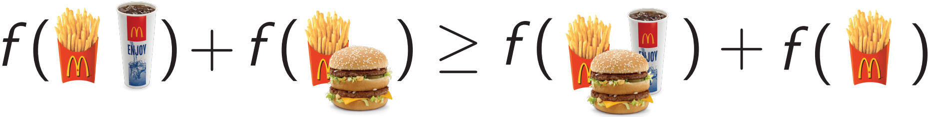


- Rearranging terms, we can see this as diminishing returns:

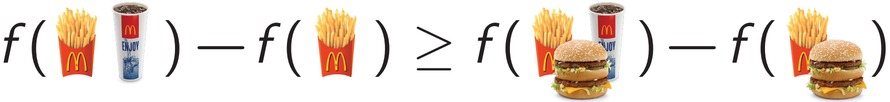


# Ex. Submodular: Consumer Costs of Living

- Consumer costs are very often submodular. For example:



- Rearranging terms, we can see this as diminishing returns:

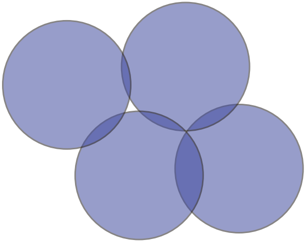


- This is very common: The additional cost of a coke is, say, free if you add it to fries and a hamburger, but when added just to an order of fries, the coke is not free.

# Area of the union of areas indexed by $A$

- Let $V$ be a set of indices, and each $v \in V$ indexes a given sub-area of some region. Let $\text{area}(v)$ be the area corresponding to item $v$.
- Let $f(S) = \bigcup_{s \in S} \text{area}(s)$ be the union of the areas indexed by elements in $\cancel{A}$. $S$
- Then $f(S)$ is submodular.

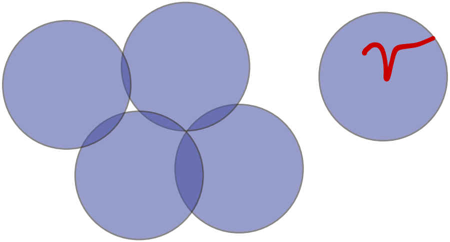
# Area of the union of areas indexed by $A$



Union of areas of elements of $A$ is given by:

$$f(A) = f(\{a_1, a_2, a_3, a_4\})$$

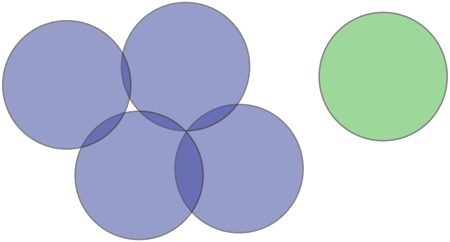
# Area of the union of areas indexed by $A$




Area of $A$ along with with $v$:

$$f(A \cup \{v\}) = f(\{a_1, a_2, a_3, a_4\} \cup \{v\})$$

# Area of the union of areas indexed by $A$

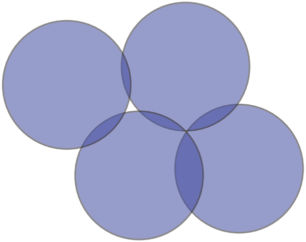


Gain (value) of $v$ in context of $A$:

$$f(A \cup \{v\}) - f(A) = f(\{v\})$$

We get full value $f(\{v\})$ in this case since the area of $v$ has no overlap with that of $A$.

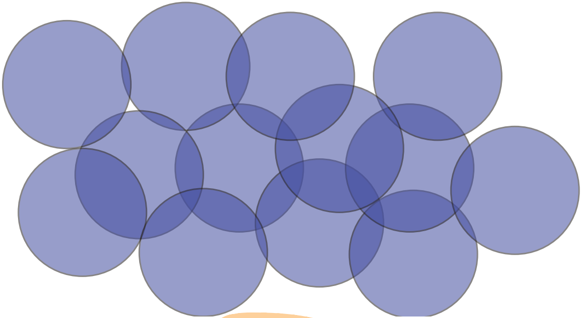
# Area of the union of areas indexed by $A$



Area of $A$ once again.

$$f(A) = f(\{a_1, a_2, a_3, a_4\})$$

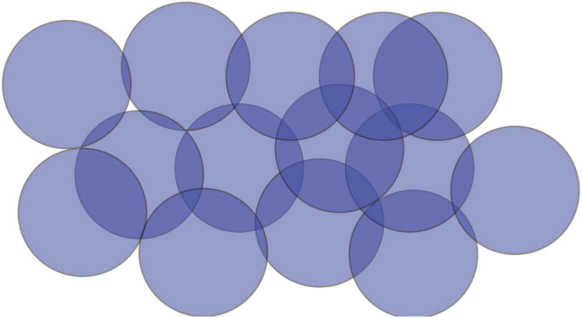
# Area of the union of areas indexed by $A$



Union of areas of elements of $B \supset A$, where $v$ is not included:

$$f(B) \text{ where } v \notin B \text{ and where } A \subseteq B$$

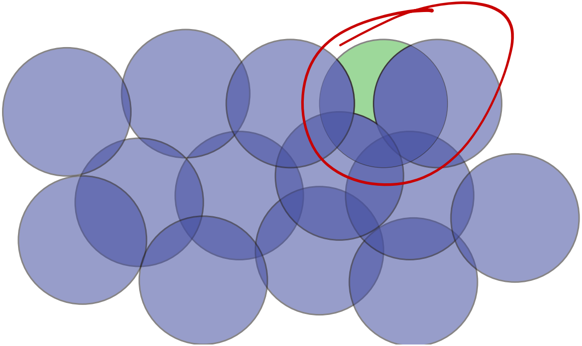
# Area of the union of areas indexed by $A$



Area of $B$ now also including $v$:

$$f(B \cup \{v\})$$

# Area of the union of areas indexed by $A$



Incremental value of $v$ in the context of $B \supset A$.

$$f(B \cup \{v\}) - f(B) < f(\{v\}) = f(A \cup \{v\}) - f(A)$$

So benefit of $v$ in the context of $A$ is greater than the benefit of $v$ in the context of $B \supseteq A$.

# Example Submodular: Entropy from Information Theory

- Entropy is submodular. Let $V$ be the index set of a set of random variables, then the function

$$f(A) = H(X_A) = -\sum_{x_A} p(x_A) \log p(x_A) \tag{2.15}$$

  is submodular.
- Proof: conditioning reduces entropy. With $A \subseteq B$ and $v \notin B$,

$$H(X_v|X_B) = H(X_{B+v}) - H(X_B) \tag{2.16}$$
$$\leq H(X_{A+v}) - H(X_A) = H(X_v|X_A) \tag{2.17}$$

# Example Submodular: Entropy from Information Theory

- Alternate Proof: Conditional mutual Information is always non-negative.
- Given $A, B, C \subseteq V$, consider conditional mutual information quantity:

$$I(X_{A \setminus B}; X_{B \setminus A} | X_{A \cap B}) = \sum_{x_{A \cup B}} p(x_{A \cup B}) \log \frac{p(x_{A \setminus B}, x_{B \setminus A} | x_{A \cap B})}{p(x_{A \setminus B} | x_{A \cap B}) p(x_{B \setminus A} | x_{A \cap B})}$$

A B $A \setminus B$ $A \cap B$ $B \setminus A$

$$= \boxed{\sum_{x_{A \cup B}} p(x_{A \cup B}) \log \frac{p(x_{A \cup B}) p(x_{A \cap B})}{p(x_A) p(x_B)} \geq 0} \tag{2.18}$$

then

$$I(X_{A \setminus B}; X_{B \setminus A} | X_{A \cap B}) \\ = H(X_A) + H(X_B) - H(X_{A \cup B}) - H(X_{A \cap B}) \geq 0 \tag{2.19}$$

so entropy satisfies

$$H(X_A) + H(X_B) \geq H(X_{A \cup B}) + H(X_{A \cap B}) \tag{2.20}$$

# Example Submodular: Mutual Information

- Also, symmetric mutual information is submodular,

$$f(A) = I(X_A; X_{V \setminus A}) = H(X_A) + H(X_{V \setminus A}) - H(X_V) \tag{2.21}$$

Note that $f(A) = H(X_A)$ and $\bar{f}(A) = H(X_{V \setminus A})$, and adding submodular functions preserves submodularity (which we will see quite soon).

# Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.

# Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.
- If $G$ is undirected, define

$$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{2.22}$$

as the edges between $X$ and $Y$.

# Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.
- If $G$ is undirected, define

$$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{2.22}$$

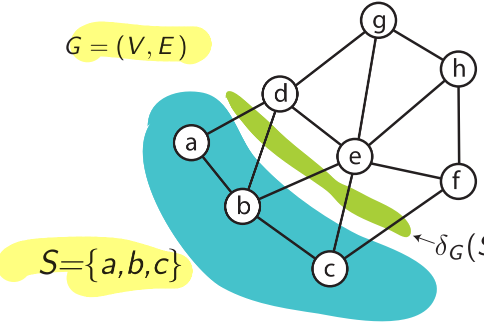
as the edges between $X$ and $Y$.
- Nodes define cuts, define the cut function $\delta(X) = E(X, V \setminus X)$.

# Undirected Graphs

- Let $G = (V, E)$ be a graph with vertices $V = V(G)$ and edges $E = E(G) \subseteq V \times V$.
- If $G$ is undirected, define

$$E(X, Y) = \{\{x, y\} \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{2.22}$$

  as the edges between $X$ and $Y$.
- Nodes define cuts, define the cut function $\delta(X) = E(X, V \setminus X)$.




$$\delta_G(S) = \{\{u, v\} \in E : u \in S, v \in V \setminus S\}.$$
$$= \{\{a,d\},\{b,d\},\{b,e\},\{c,e\},\{c,f\}\}$$

# Directed graphs, and cuts and flows

- If $G$ is directed, define

$$E^+(X,Y) \triangleq \{(x,y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{2.23}$$

as the edges directed from $X$ towards $Y$.

# Directed graphs, and cuts and flows

- If $G$ is directed, define

$$E^+(X,Y) \triangleq \{(x,y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{2.23}$$

as the edges directed from $X$ towards $Y$.

- Nodes define cuts and flows. Define edges leaving $X$ (out-flow) as

$$\delta^+(X) \triangleq E^+(X, V \setminus X) \tag{2.24}$$

and edges entering $X$ (in-flow) as

$$\delta^-(X) \triangleq E^+(V \setminus X, X) \tag{2.25}$$

# Directed graphs, and cuts and flows

- If $G$ is directed, define

$$E^+(X,Y) \triangleq \{(x,y) \in E(G) : x \in X \setminus Y, y \in Y \setminus X\} \tag{2.23}$$

as the edges directed from $X$ towards $Y$.

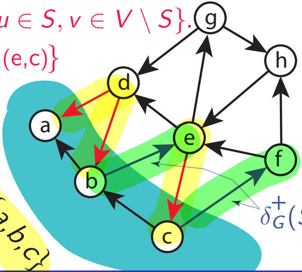
- Nodes define cuts and flows. Define edges leaving $X$ (out-flow) as

$$\delta^+(X) \triangleq E^+(X, V \setminus X) \tag{2.24}$$

and edges entering $X$ (in-flow) as

$$\delta^-(X) \triangleq E^+(V \setminus X, X) \tag{2.25}$$

$\delta^-_G(S) = \{(v,u) \in E : u \in S, v \in V \setminus S\}.$
$= \{(d,a), (d,b), (e,c)\}$




$S = \{a,b,c\}$

$\delta^+_G(S) = \{(u,v) \in E : u \in S, v \in V \setminus S\}.$
$= \{(b,e), (c,f)\}$

# The Neighbor function in undirected graphs

- Given a set $X \subseteq V$, the neighbors function of $X$ is defined as

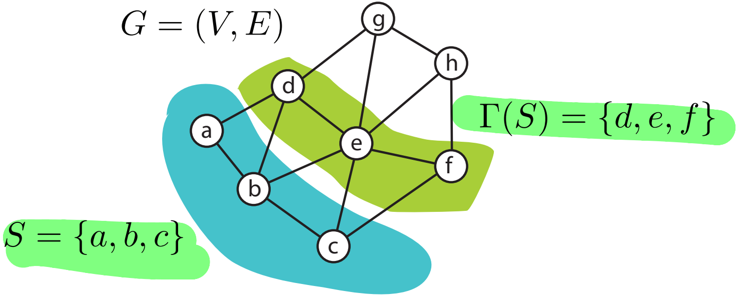
$$\Gamma(X) \triangleq \{v \in V(G) \setminus X : E(X, \{v\}) \neq \emptyset\} \tag{2.26}$$

# The Neighbor function in undirected graphs

- Given a set $X \subseteq V$, the neighbors function of $X$ is defined as

$$\Gamma(X) \triangleq \{v \in V(G) \setminus X : E(X, \{v\}) \neq \emptyset\} \quad (2.26)$$

- Example:



# Directed Cut function: property

**Lemma 2.6.1**

*For a digraph $G = (V, E)$ and any $X, Y \subseteq V$: we have*

$$|\delta^+(X)| + |\delta^+(Y)|$$
$$\geq |\delta^+(X \cap Y)| + |\delta^+(X \cup Y)| + \cancel{|E^+(X,Y)| + |E^+(Y,X)|} \quad (2.27)$$

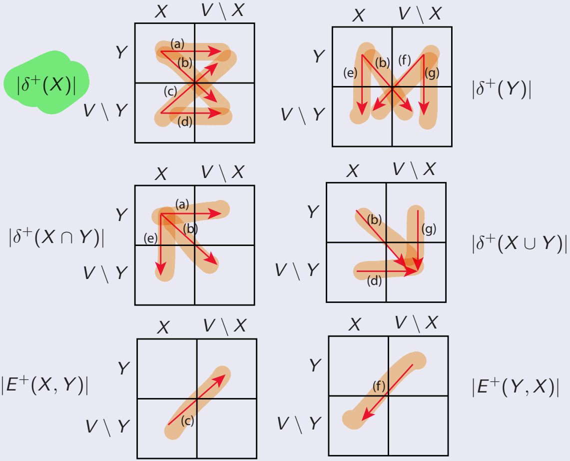
*and*

$$|\delta^-(X)| + |\delta^-(Y)|$$
$$= |\delta^-(X \cap Y)| + |\delta^-(X \cup Y)| + |E^-(X,Y)| + |E^-(Y,X)| \quad (2.28)$$

# Directed Cut function: proof of property

## Proof.

We can prove this using a simple geometric counting argument ($\delta^-(X)$ is similar)



# Directed cut/flow functions: submodular

**Lemma 2.6.2**

*For a digraph $G = (V, E)$ and any $X, Y \subseteq V$: both functions $|\delta^+(X)|$ and $|\delta^-(X)|$ are submodular.*

**Proof.**

$|E^+(X,Y)| \geq 0$ and $|E^-(X,Y)| \geq 0$. $\square$

More generally, in the non-negative weighted case, both in-flow and out-flow are submodular on subsets of the vertices.

# Undirected Cut/Flow & the Neighbor function: submodular

## Lemma 2.6.3

*For an undirected graph $G = (V, E)$ and any $X, Y \subseteq V$: we have that both the undirected cut (or flow) function $|\delta(X)|$ and the neighbor function $|\Gamma(X)|$ are submodular. I.e.,*

$$|\delta(X)| + |\delta(Y)| = |\delta(X \cap Y)| + |\delta(X \cup Y)| + 2|E(X, Y)| \tag{2.29}$$

*and*

$$|\Gamma(X)| + |\Gamma(Y)| \geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)| \tag{2.30}$$

## Proof.

- Eq. (2.29) follows from Eq. (2.27): we replace each undirected edge $\{u, v\}$ with two oppositely-directed directed edges $(u, v)$ and $(v, u)$. Then we use same counting argument.

...

# Undirected Cut/Flow & the Neighbor function: submodular

## Lemma 2.6.3

*For an undirected graph $G = (V, E)$ and any $X, Y \subseteq V$: we have that both the undirected cut (or flow) function $|\delta(X)|$ and the neighbor function $|\Gamma(X)|$ are submodular. I.e.,*

$$|\delta(X)| + |\delta(Y)| = |\delta(X \cap Y)| + |\delta(X \cup Y)| + 2|E(X, Y)| \tag{2.29}$$

*and*

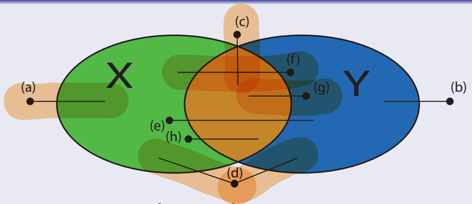
$$|\Gamma(X)| + |\Gamma(Y)| \geq |\Gamma(X \cap Y)| + |\Gamma(X \cup Y)| \tag{2.30}$$

## Proof.

- Eq. (2.29) follows from Eq. (2.27): we replace each undirected edge $\{u, v\}$ with two oppositely-directed directed edges $(u, v)$ and $(v, u)$. Then we use same counting argument.
- Eq. (2.30) follows as shown in the following page.

. . .

# cont.




Graphically, we can count and see that

$$\Gamma(X) = (a) + (c) + (f) + (g) + (d) \tag{2.31}$$

$$\Gamma(Y) = (b) + (c) + (e) + (h) + (d) \tag{2.32}$$

$$\Gamma(X \cup Y) = (a) + (b) + (c) + (d) \tag{2.33}$$

$$\Gamma(X \cap Y) = (c) + (g) + (h) \tag{2.34}$$

so

$$\begin{aligned} |\Gamma(X)| + |\Gamma(Y)| &= (a) + (b) + 2(c) + 2(d) + (e) + (f) + (g) + (h) \\ &\geq (a) + (b) + 2(c) + (d) + (g) + (h) = |\Gamma(X \cup Y)| + |\Gamma(X \cap Y)| \end{aligned} \tag{2.35}$$

# Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $\Gamma(X)$ is submodular.

# Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $\Gamma(X)$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes $u, v$ and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.

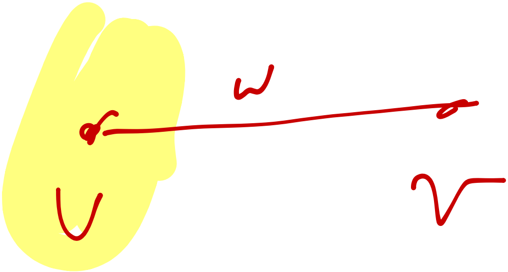
# Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $\Gamma(X)$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes $u, v$ and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut function over those two nodes: $\Gamma_{u,v}$ has valuation:

$$\Gamma_{u,v}(\emptyset) = \Gamma_{u,v}(\{u, v\}) = 0 \tag{2.36}$$

and

$$\Gamma_{u,v}(\{u\}) = \Gamma_{u,v}(\{v\}) = w \geq 0 \tag{2.37}$$



# Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $\Gamma(X)$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes $u, v$ and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut function over those two nodes: $\Gamma_{u,v}$ has valuation:

$$\Gamma_{u,v}(\emptyset) = \Gamma_{u,v}(\{u, v\}) = 0 \tag{2.36}$$

and

$$\Gamma_{u,v}(\{u\}) = \Gamma_{u,v}(\{v\}) = w \geq 0 \tag{2.37}$$

- Thus, $\Gamma_{u,v}$ is submodular since

$$\Gamma_{u,v}(\{u\}) + \Gamma_{u,v}(\{v\}) \geq \Gamma_{u,v}(\{u, v\}) + \Gamma_{u,v}(\emptyset) \tag{2.38}$$

# Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $\Gamma(X)$ is submodular.
- Define a graph $G_{uv} = (\{u,v\},\{e\},w)$ with two nodes $u,v$ and one edge $e = \{u,v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut function over those two nodes: $\Gamma_{u,v}$ has valuation:

$$\Gamma_{u,v}(\emptyset) = \Gamma_{u,v}(\{u,v\}) = 0 \tag{2.36}$$

and

$$\Gamma_{u,v}(\{u\}) = \Gamma_{u,v}(\{v\}) = w \geq 0 \tag{2.37}$$

- Thus, $\Gamma_{u,v}$ is submodular since

$$\Gamma_{u,v}(\{u\}) + \Gamma_{u,v}(\{v\}) \geq \Gamma_{u,v}(\{u,v\}) + \Gamma_{u,v}(\emptyset) \tag{2.38}$$

- General non-negative weighted graph $G = (V,E,w)$, define $\Gamma$ as:

$$\Gamma(A) = \sum_{(u,v)\in E(G)} \Gamma_{u,v}(A \cap \{u,v\}) \tag{2.39}$$

# Undirected cut/flow is submodular: alternate proof

- Another simple proof shows that $\Gamma(X)$ is submodular.
- Define a graph $G_{uv} = (\{u, v\}, \{e\}, w)$ with two nodes $u, v$ and one edge $e = \{u, v\}$ with non-negative weight $w(e) \in \mathbb{R}_+$.
- Cut function over those two nodes: $\Gamma_{u,v}$ has valuation:

$$\Gamma_{u,v}(\emptyset) = \Gamma_{u,v}(\{u, v\}) = 0 \tag{2.36}$$

and

$$\Gamma_{u,v}(\{u\}) = \Gamma_{u,v}(\{v\}) = w \geq 0 \tag{2.37}$$

- Thus, $\Gamma_{u,v}$ is submodular since

$$\Gamma_{u,v}(\{u\}) + \Gamma_{u,v}(\{v\}) \geq \Gamma_{u,v}(\{u, v\}) + \Gamma_{u,v}(\emptyset) \tag{2.38}$$

- General non-negative weighted graph $G = (V, E, w)$, define $\Gamma$ as:

$$\Gamma(A) = \sum_{(u,v) \in E(G)} \Gamma_{u,v}(A \cap \{u, v\}) \tag{2.39}$$

- This is easily shown to be submodular using properties we will soon see (namely, submodularity closed under summation and restriction).

# Undirected Neighbor functions

Therefore, the undirected cut function $|\delta(A)|$ and the neighbor function $|\Gamma(A)|$ of a graph $G$ are both submodular.

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$.

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function.

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function. If you had to guess, is this always the case?

# Other graph functions that are submodular/supermodular

These come from Narayanan's book 1997. Let $G$ be an undirected graph.

- Let $V(X)$ be the vertices adjacent to some edge in $X \subseteq E(G)$, then $|V(X)|$ (the vertex function) is submodular.
- Let $E(S)$ be the edges with both vertices in $S \subseteq V(G)$. Then $|E(S)|$ (the interior edge function) is supermodular.
- Let $I(S)$ be the edges with at least one vertex in $S \subseteq V(G)$. Then $|I(S)|$ (the incidence function) is submodular.
- Recall $|\delta(S)|$, is the set size of edges with exactly one vertex in $S \subseteq V(G)$ is submodular (cut size function). Thus, we have $I(S) = E(S) \cup \delta(S)$ and $E(S) \cap \delta(S) = \emptyset$, and thus that $|I(S)| = |E(S)| + |\delta(S)|$. So we can get a submodular function by summing a submodular and a supermodular function. If you had to guess, is this always the case?
- Consider $f(A) = |\delta^+(A)| - |\delta^+(V \setminus A)|$. Guess, submodular, supermodular, modular, or neither? Exercise: determine which one and prove it.

# Other graph functions that are submodular/supermodular

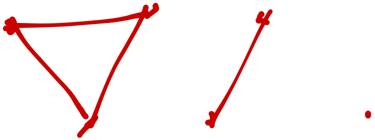
- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.

# Other graph functions that are submodular/supermodular

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.

# Other graph functions that are submodular/supermodular

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.



# Other graph functions that are submodular/supermodular

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- Then $c(A)$ is supermodular. Intuition why: adding an edge to a larger context of edges is "more" (i.e., no less) able to bridge separate connected components than when the edge is added to a smaller context of edges.

# Other graph functions that are submodular/supermodular

- Recall, $f : 2^V \to \mathbb{R}$ is submodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Hence, if $f : 2^V \to \mathbb{R}$ is supermodular, then so is $\bar{f} : 2^V \to \mathbb{R}$ defined as $\bar{f}(S) = f(V \setminus S)$.
- Given a graph $G = (V, E)$, for each $A \subseteq E(G)$, let $c(A)$ denote the number of connected components of the (spanning) subgraph $(V(G), A)$, with $c : 2^E \to \mathbb{R}_+$.
- Then $c(A)$ is supermodular. Intuition why: adding an edge to a larger context of edges is "more" (i.e., no less) able to bridge separate connected components than when the edge is added to a smaller context of edges.
- $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, and hence is also supermodular.

# Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, is supermodular.

# Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).

# Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set $A$ and shatter the graph into many connected components, then the graph is weak.

# Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set $A$ and shatter the graph into many connected components, then the graph is weak.
- An attacker wishes to choose a small number of edges (since it is cheap) to shatter the graph into as many components as possible.

# Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set $A$ and shatter the graph into many connected components, then the graph is weak.
- An attacker wishes to choose a small number of edges (since it is cheap) to shatter the graph into as many components as possible.
- Let $G = (V, E, w)$ with $w : E \to \mathbb{R}+$ be a weighted graph with non-negative weights.

# Graph Strength

- So $\bar{c}(A) = c(E \setminus A)$ is the number of connected components in $G$ when we remove $A$, is supermodular.
- Maximizing $\bar{c}(A)$ might seem as a goal for a network attacker — many connected components means that many points in the network have lost connectivity to many other points (unprotected network).
- If we can remove a small set $A$ and shatter the graph into many connected components, then the graph is weak.
- An attacker wishes to choose a small number of edges (since it is cheap) to shatter the graph into as many components as possible.
- Let $G = (V, E, w)$ with $w : E \to \mathbb{R}+$ be a weighted graph with non-negative weights.
- For $(u, v) = e \in E$, let $w(e)$ be a measure of the strength of the connection between vertices $u$ and $v$ (strength meaning the difficulty of cutting the edge $e$).

# Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function

$$w(A) = \sum_{e \in A} w_e \tag{2.40}$$

so that $w(E(G[S]))$ is the "internal strength" of the vertex set $S$.
*Notation: $S$ is a set of nodes, $G[S]$ is the vertex-induced subgraph of $G$ induced by vertices $S$, $E(G[S])$ are the edges contained within this induced subgraph, and $w(E(G[S]))$ is the weight of these edges.*

# Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function

$$w(A) = \sum_{e \in A} w_e \tag{2.40}$$

so that $w(E(G[S]))$ is the "internal strength" of the vertex set $S$.

- Suppose removing $A$ shatters $G$ into a graph with $\bar{c}(A) > 1$ components — then $w(A)/(\bar{c}(A) - 1)$ is like the "effort per achieved component" for a network attacker.

# Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function

$$w(A) = \sum_{e \in A} w_e \tag{2.40}$$

so that $w(E(G[S]))$ is the "internal strength" of the vertex set $S$.

- Suppose removing $A$ shatters $G$ into a graph with $\bar{c}(A) > 1$ components — then $w(A)/(\bar{c}(A) - 1)$ is like the "effort per achieved component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G) : \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{2.41}$$

# Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function

$$w(A) = \sum_{e \in A} w_e \tag{2.40}$$

so that $w(E(G[S]))$ is the "internal strength" of the vertex set $S$.

- Suppose removing $A$ shatters $G$ into a graph with $\bar{c}(A) > 1$ components — then $w(A)/(\bar{c}(A) - 1)$ is like the "effort per achieved component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G): \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{2.41}$$

- Graph strength is like the minimum effort per component. An attacker would use the argument of the min to choose which edges to attack. A network designer would maximize, over $G$ and/or $w$, the graph strength, $\textit{strength}(G, w)$.

# Graph Strength

- Then $w(A)$ for $A \subseteq E$ is a modular function

$$w(A) = \sum_{e \in A} w_e \tag{2.40}$$

so that $w(E(G[S]))$ is the "internal strength" of the vertex set $S$.

- Suppose removing $A$ shatters $G$ into a graph with $\bar{c}(A) > 1$ components — then $w(A)/(\bar{c}(A) - 1)$ is like the "effort per achieved component" for a network attacker.
- A form of graph strength can then be defined as the following:

$$\textit{strength}(G, w) = \min_{A \subseteq E(G) : \bar{c}(A) > 1} \frac{w(A)}{\bar{c}(A) - 1} \tag{2.41}$$

- Graph strength is like the minimum effort per component. An attacker would use the argument of the min to choose which edges to attack. A network designer would maximize, over $G$ and/or $w$, the graph strength, *strength*$(G, w)$.
- Since submodularity, problems have strongly-poly-time solutions.

# Matrix Rank functions

- Let $V$ be an index set of a set of vectors in $\mathbb{R}^n$ m for some $n$.

# Matrix Rank functions

- Let $V$ be an index set of a set of vectors in $\mathbb{R}^n$ for some $n$.
- For a given set $\{v, v_1, v_2, \dots, v_k\}$, it might or might not be possible to find $(\alpha_i)_i$ such that:

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{2.42}$$

If not, then $x_v$ is linearly independent of $x_{v_1}, \dots, x_{v_k}$.

# Matrix Rank functions

- Let $V$ be an index set of a set of vectors in $\mathbb{R}^n$ for some $n$.
- For a given set $\{v, v_1, v_2, \dots, v_k\}$, it might or might not be possible to find $(\alpha_i)_i$ such that:

$$x_v = \sum_{i=1}^{k} \alpha_i x_{v_i} \tag{2.42}$$

  If not, then $x_v$ is linearly independent of $x_{v_1}, \dots, x_{v_k}$.
- Let $r(S)$ for $S \subseteq V$ be the rank of the set of vectors $S$. Then $r(\cdot)$ is a submodular function, and in fact is called a matric matroid rank function.

# Example: Rank function of a matrix

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \ldots, x_m)$. There are $m$ length-$n$ column vectors $\{x_i\}_i$

Skip matrix rank example

# Example: Rank function of a matrix

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \ldots, x_m)$. There are $m$ length-$n$ column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \ldots, m\}$ be the set of column vector indices.

Skip matrix rank example

# Example: Rank function of a matrix

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \dots, x_m)$. There are $m$ length-$n$ column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \dots, m\}$ be the set of column vector indices.
- For any $A \subseteq V$, let $r(A)$ be the rank of the column vectors indexed by $A$.

Skip matrix rank example

# Example: Rank function of a matrix

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \dots, x_m)$. There are $m$ length-$n$ column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \dots, m\}$ be the set of column vector indices.
- For any $A \subseteq V$, let $r(A)$ be the rank of the column vectors indexed by $A$.
- $r(A)$ is the dimensionality of the vector space spanned by the set of vectors $\{x_a\}_{a \in A}$.

Skip matrix rank example

# Example: Rank function of a matrix

- Given $n \times m$ matrix $\mathbf{X} = (x_1, x_2, \dots, x_m)$. There are $m$ length-$n$ column vectors $\{x_i\}_i$
- Let $V = \{1, 2, \dots, m\}$ be the set of column vector indices.
- For any $A \subseteq V$, let $r(A)$ be the rank of the column vectors indexed by $A$.
- $r(A)$ is the dimensionality of the vector space spanned by the set of vectors $\{x_a\}_{a \in A}$.
- Thus, $r(V)$ is the rank of the matrix $\mathbf{X}$.

Skip matrix rank example

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c c} & \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array}
= \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\
\begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c c} & \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}
\begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array}
=
\begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c|cccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \hline 1 & 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 2 & 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 3 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{array} = \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1,2,3,4,5,6,7,8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1,2,3\}$, $B = \{3,4,5\}$, $C = \{6,7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{6, 7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1$ $< r(C) = 2.$

# Example: Rank function of a matrix

Consider the following $4 \times 8$ matrix, so $V = \{1,2,3,4,5,6,7,8\}$.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} 0 & 2 & 2 & 3 & 0 & 1 & 3 & 1 \\ 0 & 3 & 0 & 4 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix} \end{array} = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array} \\ \begin{pmatrix} | & | & | & | & | & | & | & | \\ x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \\ | & | & | & | & | & | & | & | \end{pmatrix} \end{array}$$

- Let $A = \{1,2,3\}$, $B = \{3,4,5\}$, $C = \{6,7\}$, $A_r = \{1\}$, $B_r = \{5\}$.
- Then $r(A) = 3$, $r(B) = 3$, $r(C) = 2$.
- $r(A \cup C) = 3$, $r(B \cup C) = 3$.
- $r(A \cup A_r) = 3$, $r(B \cup B_r) = 3$, $r(A \cup B_r) = 4$, $r(B \cup A_r) = 4$.
- $r(A \cup B) = 4$, $r(A \cap B) = 1 < r(C) = 2$.
- $6 = r(A) + r(B) > r(A \cup B) + r(A \cap B) = 5$

# Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.

# Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.

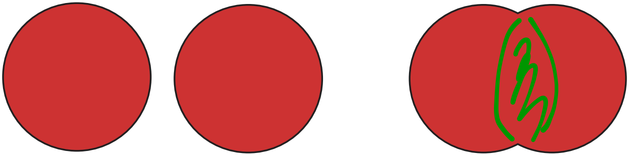
# Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.
- In Venn diagram, Let area correspond to dimensions spanned by vectors indexed by a set. Hence, $r(A)$ can be viewed as an area.

# Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.
- In Venn diagram, Let area correspond to dimensions spanned by vectors indexed by a set. Hence, $r(A)$ can be viewed as an area.

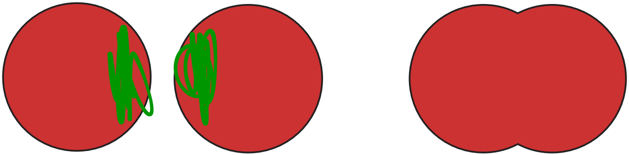
$$r(A) \quad + \quad r(B) \qquad \geq \qquad r(A \cup B)$$



# Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.
- In Venn diagram, Let area correspond to dimensions spanned by vectors indexed by a set. Hence, $r(A)$ can be viewed as an area.

$$r(A) \quad + \quad r(B) \qquad \geq \qquad r(A \cup B)$$

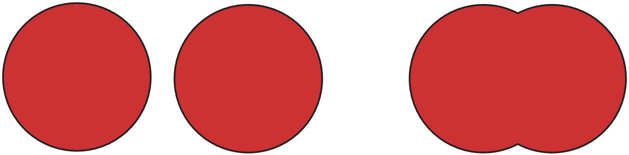


- If some of the dimensions spanned by $A$ overlap some of the dimensions spanned by $B$ (i.e., if $\exists$ common span), then that area is counted twice in $r(A) + r(B)$, so the inequality will be strict.

# Rank function of a matrix

- Let $A, B \subseteq V$ be two subsets of column indices.
- The rank of the two sets unioned together $A \cup B$ is no more than the sum of the two individual ranks.
- In Venn diagram, Let area correspond to dimensions spanned by vectors indexed by a set. Hence, $r(A)$ can be viewed as an area.

$$r(A) \quad + \quad r(B) \qquad \geq \qquad r(A \cup B)$$



- If some of the dimensions spanned by $A$ overlap some of the dimensions spanned by $B$ (i.e., if $\exists$ common span), then that area is counted twice in $r(A) + r(B)$, so the inequality will be strict.
- Any function where the above inequality is true for all $A, B \subseteq V$ is called subadditive.

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$.

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$.
- Let $B_r$ index vectors spanning dimensions spanned by $B$ but not $A$.

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$.
- Let $B_r$ index vectors spanning dimensions spanned by $B$ but not $A$.
- Then, $r(A) = r(C) + r(A_r)$

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$.
- Let $B_r$ index vectors spanning dimensions spanned by $B$ but not $A$.
- Then, $r(A) = r(C) + r(A_r)$
- Similarly, $r(B) = r(C) + r(B_r)$.

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$.
- Let $B_r$ index vectors spanning dimensions spanned by $B$ but not $A$.
- Then, $r(A) = r(C) + r(A_r)$
- Similarly, $r(B) = r(C) + r(B_r)$.
- Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{2.43}$$

# Rank functions of a matrix

- Vectors $A$ and $B$ have a (possibly empty) common span and two (possibly empty) non-common residual spans.
- Let $C$ index vectors spanning dimensions common to $A$ and $B$.
- Let $A_r$ index vectors spanning dimensions spanned by $A$ but not $B$.
- Let $B_r$ index vectors spanning dimensions spanned by $B$ but not $A$.
- Then, $r(A) = r(C) + r(A_r)$
- Similarly, $r(B) = r(C) + r(B_r)$.
- Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r). \tag{2.43}$$

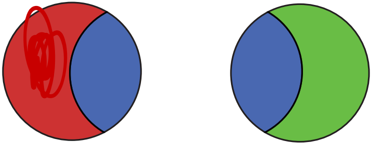
- But $r(A \cup B)$ counts the dimensions spanned by $C$ only once.

$$r(A \cup B) = r(A_r) + r(C) + r(B_r) \tag{2.44}$$

# Rank functions of a matrix

- Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,

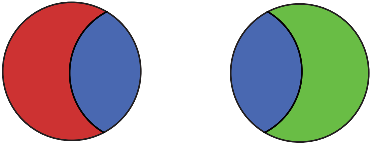
$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$



# Rank functions of a matrix

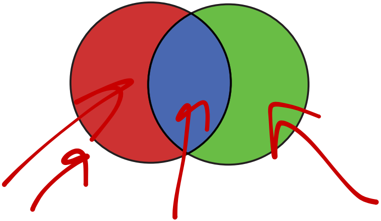
- Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$



- But $r(A \cup B)$ counts the dimensions spanned by $C$ only once.

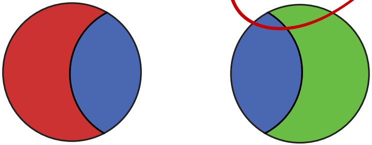
$$r(A \cup B) = r(A_r) + r(C) + r(B_r)$$



# Rank functions of a matrix

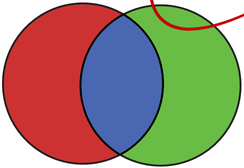
- Then $r(A) + r(B)$ counts the dimensions spanned by $C$ twice, i.e.,

$$r(A) + r(B) = r(A_r) + 2r(C) + r(B_r)$$



- But $r(A \cup B)$ counts the dimensions spanned by $C$ only once.

$$r(A \cup B) = r(A_r) + r(C) + r(B_r)$$

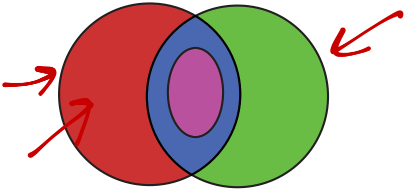


- Thus, we have subadditivity: $r(A) + r(B) \geq r(A \cup B)$. Can we add more to the r.h.s. and still have an inequality? Yes.

# Rank function of a matrix

- Note, $r(A \cap B) \leq r(C)$. Why? Vectors indexed by $A \cap B$ (i.e., the common index set) span no more than the dimensions commonly spanned by $A$ and $B$ (namely, those spanned by the professed $C$).

$$r(C) \geq r(A \cap B)$$

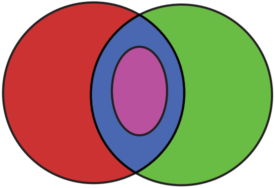


In short:

# Rank function of a matrix

- Note, $r(A \cap B) \leq r(C)$. Why? Vectors indexed by $A \cap B$ (i.e., the common index set) span no more than the dimensions commonly spanned by $A$ and $B$ (namely, those spanned by the professed $C$).

$$r(C) \geq r(A \cap B)$$



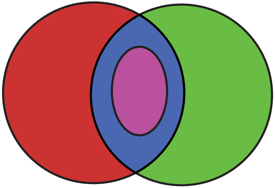
In short:

- Common span (blue) is "more" (no less) than span of common index (magenta).

# Rank function of a matrix

- Note, $r(A \cap B) \leq r(C)$. Why? Vectors indexed by $A \cap B$ (i.e., the common index set) span no more than the dimensions commonly spanned by $A$ and $B$ (namely, those spanned by the professed $C$).

$$r(C) \geq r(A \cap B)$$

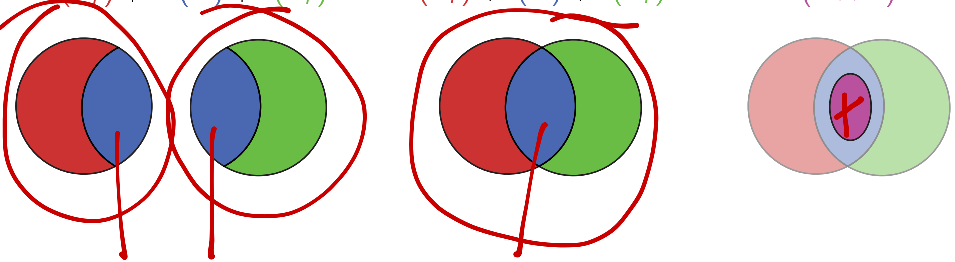


In short:

- Common span (blue) is "more" (no less) than span of common index (magenta).
- More generally, common information (blue) is "more" (no less) than information within common index (magenta).

# The Venn and Art of Submodularity

$$\underbrace{r(A) + r(B)}_{= r(A_r) + 2r(C) + r(B_r)} \geq \underbrace{r(A \cup B)}_{= r(A_r) + r(C) + r(B_r)} + \underbrace{r(A \cap B)}_{= r(A \cap B)}$$



# Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let (i.e., each $s \in S$ is a subspace of dimension $\geq 1$).

# Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let (i.e., each $s \in S$ is a subspace of dimension $\geq 1$).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in $X$.

# Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let (i.e., each $s \in S$ is a subspace of dimension $\geq 1$).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in $X$.
- We can think of $S$ as a set of sets of vectors from the previous example, and for each $s \in S$, let $X_s$ being a set of vector indices.

# Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let (i.e., each $s \in S$ is a subspace of dimension $\geq 1$).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in $X$.
- We can think of $S$ as a set of sets of vectors from the previous example, and for each $s \in S$, let $X_s$ being a set of vector indices.
- Then, defining $f : 2^S \to \mathbb{R}_+$ as follows,

$$f(X) = r(\cup_{s \in S} X_s) \tag{2.45}$$

we have that $f$ is submodular, and is known to be a polymatroid rank function.

# Polymatroid rank function

- Let $S$ be a set of subspaces of a linear space and let (i.e., each $s \in S$ is a subspace of dimension $\geq 1$).
- For each $X \subseteq S$, let $f(X)$ denote the dimensionality of the linear subspace spanned by the subspaces in $X$.
- We can think of $S$ as a set of sets of vectors from the previous example, and for each $s \in S$, let $X_s$ being a set of vector indices.
- Then, defining $f : 2^S \to \mathbb{R}_+$ as follows,

$$f(X) = r(\cup_{s \in S} X_s) \tag{2.45}$$

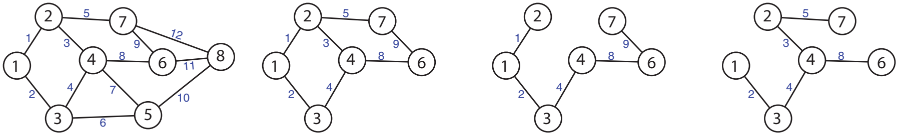
  we have that $f$ is submodular, and is known to be a polymatroid rank function.
- In general (as we will see) polymatroid rank functions are submodular, normalized $f(\emptyset) = 0$, and monotone non-decreasing ($f(A) \leq f(B)$ whenever $A \subseteq B$).

# Spanning trees

- Let $E$ be a set of edges of some graph $G = (V, E)$, and let $r(S)$ for $S \subseteq E$ be the maximum size (in terms of number of edges) spanning forest in the vertex-induced graph induced by edges adjacent to $S$.

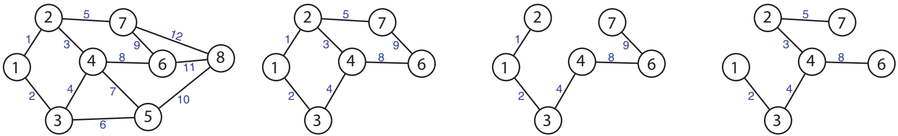
# Spanning trees

- Let $E$ be a set of edges of some graph $G = (V, E)$, and let $r(S)$ for $S \subseteq E$ be the maximum size (in terms of number of edges) spanning forest in the vertex-induced graph induced by edges adjacent to $S$.
- Example: Given $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \ldots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\}$. Two spanning trees have the same edge count (the rank of $S$).



# Spanning trees

- Let $E$ be a set of edges of some graph $G = (V, E)$, and let $r(S)$ for $S \subseteq E$ be the maximum size (in terms of number of edges) spanning forest in the vertex-induced graph induced by edges adjacent to $S$.
- Example: Given $G = (V, E)$, $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $E = \{1, 2, \ldots, 12\}$. $S = \{1, 2, 3, 4, 5, 8, 9\}$. Two spanning trees have the same edge count (the rank of $S$).




- Then $r(S)$ is submodular, and is another matrix rank function corresponding to the incidence matrix of the graph.

# Supply Side Economies of scale

- What is a good model of the cost of manufacturing a set of items?
- Let $V$ be a set of possible items that a company might possibly wish to manufacture, and let $f(S)$ for $S \subseteq V$ be the cost to that company to manufacture subset $S$.
- Ex: $V$ might be colors of paint in a paint manufacturer: green, red, blue, yellow, white, etc.
- Producing green when you are already producing yellow and blue is probably cheaper than if you were only producing some other colors.

$$f(\text{green}, \text{blue}, \text{yellow}) - f(\text{blue}, \text{yellow}) <= f(\text{green}, \text{blue}) - f(\text{blue}) \tag{2.1}$$

- So a submodular function would be a good model.

# A model of Influence in Social Networks

- Given a graph $G = (V, E)$, each $v \in V$ corresponds to a person, to each $v$ we have an activation function $f_v : 2^V \to [0, 1]$ dependent only on its neighbors. I.e., $f_v(A) = f_v(A \cap \Gamma(v))$.
- Goal - Viral Marketing: find a small subset $S \subseteq V$ of individuals to directly influence, and thus indirectly influence the greatest number of possible other individuals (via the social network $G$).
- We define a function $f : 2^V \to \mathbb{Z}^+$ that models the ultimate influence of an initial set $S$ of nodes based on the following iterative process: At each step, a given set of nodes $S$ are activated, and we activate new nodes $v \in V \setminus S$ if $f_v(S) \geq U[0, 1]$ (where $U[0, 1]$ is a uniform random number between 0 and 1).
- It can be shown that for many $f_v$ (including simple linear functions, and where $f_v$ is submodular itself) that $f$ is submodular.

# The value of a friend

- Let $V$ be a group of individuals. How valuable to you is a given friend $v \in V$?
- It depends on how many friends you have.
- Given a group of friends $S \subseteq V$, can you valuate them with a function $f(S)$ an how?
- Let $f(S)$ be the value of the set of friends $S$. Is submodular or supermodular a good model?

# Information and Summarization

- Let $V$ be a set of information containing elements ($V$ might say be either words, sentences, documents, web pages, or blogs, each $v \in V$ is one element, so $v$ might be a word, a sentence, a document, etc.). The total amount of information in $V$ is measure by a a function $f(V)$, and any given subset $S \subseteq V$ measures the amount of information in $S$, given by $f(S)$.
- How informative is any given item $v$ in different sized contexts? Any such real-world information function would exhibit diminishing returns, i.e., the value of $v$ decreases when it is considered in a larger context.
- So a submodular function would likely be a good model.

# Submodular Polyhedra

- Submodular functions have associated polyhedra with nice properties: when a set of constraints in a linear program is a submodular polyhedron, a simple greedy algorithm can find the optimal solution even though the polyhedron is formed via an exponential number of constraints.

$$P_f = \left\{x \in \mathbb{R}^E : x(S) \le f(S), \forall S \subseteq E\right\} \tag{2.46}$$

$$P_f^+ = P_f \cap \left\{x \in \mathbb{R}^E : x \ge 0\right\} \tag{2.47}$$

$$B_f = P_f \cap \left\{x \in \mathbb{R}^E : x(E) = f(E)\right\} \tag{2.48}$$

- The linear programming problem is to, given $c \in \mathbb{R}^E$, compute:

$$\tilde{f}(c) \triangleq \max\left\{c^T x : x \in P_f\right\} \tag{2.49}$$

- This can be solved using the greedy algorithm! Moreover, $\tilde{f}(c)$ computed using greedy is convex if and only of $f$ is submodular (we will go into this in some detail this quarter).

# Ground set: $E$ or $V$?

Submodular functions are functions defined on subsets of some finite set, called the ground set .

- It is common in the literature to use either $E$ or $V$ as the ground set.

# Ground set: $E$ or $V$?

Submodular functions are functions defined on subsets of some finite set, called the ground set .

- It is common in the literature to use either $E$ or $V$ as the ground set.
- We will follow this inconsistency in the literature and will inconsistently use either $E$ or $V$ as our ground set (hopefully not in the same equation, if so, please point this out).

# Notation $\mathbb{R}^E$

What does $x \in \mathbb{R}^E$ mean?

$$\mathbb{R}^E = \{x = (x_j \in \mathbb{R} : j \in E)\} \tag{2.50}$$

$$\mathbb{R}_+^E = \{x = (x_j : j \in E) : x \geq 0\} \tag{2.51}$$

Any vector $x \in \mathbb{R}^E$ can be treated as a normalized modular function, and vice verse. That is

$$x(A) = \sum_{a \in A} x_a \tag{2.52}$$

Note that $x$ is said to be normalized since $x(\emptyset) = 0$.

# Other Notation: indicator vectors of sets

Given an $A \subseteq E$, define the vector $\mathbf{1}_A \in \mathbb{R}_+^E$ to be

$$\mathbf{1}_A(j) = \begin{cases} 1 & \text{if } j \in A; \\ 0 & \text{if } j \notin A \end{cases} \tag{2.53}$$

Sometimes this will be written as $\chi_A \equiv \mathbf{1}_A$.

# Other Notation: indicator vectors of sets

Given an $A \subseteq E$, define the vector $\mathbf{1}_A \in \mathbb{R}_+^E$ to be

$$\mathbf{1}_A(j) = \begin{cases} 1 & \text{if } j \in A; \\ 0 & \text{if } j \notin A \end{cases} \tag{2.53}$$

Sometimes this will be written as $\chi_A \equiv \mathbf{1}_A$.
Thus, given modular function $x \in \mathbb{R}^E$, we can write $x(A)$ in a variety of ways, i.e.,

$$x(A) = x \cdot \mathbf{1}_A = \sum_{i \in A} x(i) \tag{2.54}$$

# Other Notation: singletons and sets

When $A$ is a set and $k$ is a singleton (i.e., a single item), the union is properly written as $A \cup \{k\}$, but sometimes I will write just $A + k$.

# General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).

# General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \rightarrow S$.

# General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).

# General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Similarly, $2^E$ is the set of all functions from $E$ to "two" — in this case, we really mean $2 \equiv \{0, 1\}$, so $2^E$ is shorthand for $\{0, 1\}^V$

# General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Similarly, $2^E$ is the set of all functions from $E$ to "two" — in this case, we really mean $2 \equiv \{0, 1\}$, so $2^E$ is shorthand for $\{0, 1\}^V$ — hence, $2^E$ is the set of all functions that map from elements of $E$ to $\{0, 1\}$, equivalent to all binary vectors with elements indexed by elements of $E$, equivalent to subsets of $E$. Hence, if $A \in 2^E$ then $A \subseteq E$.

# General notation: what does $S^T$ mean when $S$ and $T$ are arbitrary sets

- Let $S$ and $T$ be two arbitrary sets (either of which could be countable, or uncountable).
- We define the notation $S^T$ to be the set of all functions that map from $T$ to $S$. That is, if $f \in S^T$, then $f : T \to S$.
- Hence, given a finite set $E$, $\mathbb{R}^E$ is the set of all functions that map from elements of $E$ to the reals $\mathbb{R}$, and such functions are identical to a vector in a vector space with axes labeled as elements of $E$ (i.e., if $m \in \mathbb{R}^E$, then for all $e \in E$, $m(e) \in \mathbb{R}$).
- Similarly, $2^E$ is the set of all functions from $E$ to "two" — in this case, we really mean $2 \equiv \{0, 1\}$, so $2^E$ is shorthand for $\{0, 1\}^V$ — hence, $2^E$ is the set of all functions that map from elements of $E$ to $\{0, 1\}$, equivalent to all binary vectors with elements indexed by elements of $E$, equivalent to subsets of $E$. Hence, if $A \in 2^E$ then $A \subseteq E$. What might $3^E$ mean?

# Summing Submodular Functions

Given $E$, let $f_1, f_2 : 2^E \to \mathbb{R}$ be two submodular functions. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{2.55}$$

is submodular.

# Summing Submodular Functions

Given $E$, let $f_1, f_2 : 2^E \to \mathbb{R}$ be two submodular functions. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{2.55}$$

is submodular. This follows easily since

$$\begin{aligned} f(A) + f(B) &= f_1(A) + f_2(A) + f_1(B) + f_2(B) && (2.56) \\ &\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) && (2.57) \\ &= f(A \cup B) + f(A \cap B). && (2.58) \end{aligned}$$

I.e., it holds for each component of $f$ in each term in the inequality.

# Summing Submodular Functions

Given $E$, let $f_1, f_2 : 2^E \to \mathbb{R}$ be two submodular functions. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) + f_2(A) \tag{2.55}$$

is submodular.This follows easily since

$$\begin{aligned} f(A) + f(B) &= f_1(A) + f_2(A) + f_1(B) + f_2(B) && (2.56) \\ &\geq f_1(A \cup B) + f_2(A \cup B) + f_1(A \cap B) + f_2(A \cap B) && (2.57) \\ &= f(A \cup B) + f(A \cap B). && (2.58) \end{aligned}$$

I.e., it holds for each component of $f$ in each term in the inequality. In fact, any conic combination (i.e., non-negative linear combination) of submodular functions is submodular, as in $f(A) = \alpha_1 f_1(A) + \alpha_2 f_2(A)$ for $\alpha_1, \alpha_2 \geq 0$.

# Summing Submodular and Modular Functions

Given $E$, let $f_1, m : 2^E \to \mathbb{R}$ be a submodular and a modular function.

# Summing Submodular and Modular Functions

Given $E$, let $f_1, m : 2^E \to \mathbb{R}$ be a submodular and a modular function. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{2.59}$$

is submodular (as is $f(A) = f_1(A) + m(A)$).

# Summing Submodular and Modular Functions

Given $E$, let $f_1, m : 2^E \to \mathbb{R}$ be a submodular and a modular function. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{2.59}$$

is submodular (as is $f(A) = f_1(A) + m(A)$). This follows easily since

$$f(A) + f(B) = f_1(A) - m(A) + f_1(B) - m(B) \tag{2.60}$$

$$\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) \tag{2.61}$$

$$= f(A \cup B) + f(A \cap B). \tag{2.62}$$

# Summing Submodular and Modular Functions

Given $E$, let $f_1, m : 2^E \to \mathbb{R}$ be a submodular and a modular function. Then

$$f : 2^E \to \mathbb{R} \text{ with } f(A) = f_1(A) - m(A) \tag{2.59}$$

is submodular (as is $f(A) = f_1(A) + m(A)$). This follows easily since

$$f(A) + f(B) = f_1(A) - m(A) + f_1(B) - m(B) \tag{2.60}$$

$$\geq f_1(A \cup B) - m(A \cup B) + f_1(A \cap B) - m(A \cap B) \tag{2.61}$$

$$= f(A \cup B) + f(A \cap B). \tag{2.62}$$

That is, the modular component with $m(A) + m(B) = m(A \cup B) + m(A \cap B)$ never destroys the inequality. Note of course that if $m$ is modular than so is $-m$.

# Restricting Submodular Functions

Given $E$, let $f : 2^E \to \mathbb{R}$ be a submodular functions. And let $S \subseteq E$ be an arbitrary fixed set. Then

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{2.63}$$

is submodular.

# Restricting Submodular Functions

Given $E$, let $f : 2^E \to \mathbb{R}$ be a submodular functions. And let $S \subseteq E$ be an arbitrary fixed set. Then

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{2.63}$$

is submodular.

**Proof.**

$\square$

# Restricting Submodular Functions

Given $E$, let $f : 2^E \to \mathbb{R}$ be a submodular functions. And let $S \subseteq E$ be an arbitrary fixed set. Then

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{2.63}$$

is submodular.

**Proof.**

Given $A \subseteq B \subseteq E \setminus v$, consider

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{2.64}$$

□

# Restricting Submodular Functions

Given $E$, let $f : 2^E \to \mathbb{R}$ be a submodular functions. And let $S \subseteq E$ be an arbitrary fixed set. Then

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{2.63}$$

is submodular.

**Proof.**

Given $A \subseteq B \subseteq E \setminus v$, consider

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{2.64}$$

If $v \notin S$, then both differences on each size are zero.

□

# Restricting Submodular Functions

Given $E$, let $f : 2^E \to \mathbb{R}$ be a submodular functions. And let $S \subseteq E$ be an arbitrary fixed set. Then

$$f' : 2^E \to \mathbb{R} \text{ with } f'(A) = f(A \cap S) \tag{2.63}$$

is submodular.

**Proof.**

Given $A \subseteq B \subseteq E \setminus v$, consider

$$f((A + v) \cap S) - f(A \cap S) \geq f((B + v) \cap S) - f(B \cap S) \tag{2.64}$$

If $v \notin S$, then both differences on each size are zero. If $v \in S$, then we can consider this

$$f(A' + v) - f(A') \geq f(B' + v) - f(B') \tag{2.65}$$

with $A' = A \cap S$ and $B' = B \cap S$. Since $A' \subseteq B'$, this holds due to submodularity of $f$. $\square$

# Summing Restricted Submodular Functions

Given $V$, let $f_1, f_2 : 2^V \to \mathbb{R}$ be two submodular functions and let $S_1, S_2$ be two arbitrary fixed sets. Then

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \tag{2.66}$$

is submodular. This follows easily from the preceding two results.

# Summing Restricted Submodular Functions

Given $V$, let $f_1, f_2 : 2^V \to \mathbb{R}$ be two submodular functions and let $S_1, S_2$ be two arbitrary fixed sets. Then

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = f_1(A \cap S_1) + f_2(A \cap S_2) \tag{2.66}$$

is submodular. This follows easily from the preceding two results.

Given $V$, let $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$ be a set of subsets of $V$, and for each $C \in \mathcal{C}$, let $f_C : 2^V \to \mathbb{R}$ be a submodular function. Then

$$f : 2^V \to \mathbb{R} \text{ with } f(A) = \sum_{C \in \mathcal{C}} f_C(A \cap C) \tag{2.67}$$

is submodular. This property is critical for image processing and graphical models. For example, let $\mathcal{C}$ be all pairs of the form $\{\{u, v\} : u, v \in V\}$, or let it be all pairs corresponding to the edges of some undirected graphical model. We plan to revisit this topic later in the term.

# Max - normalized

Given $V$, let $c \in \mathbb{R}_+^V$ be a given fixed vector. Then $f : 2^V \to \mathbb{R}_+$, where

$$f(A) = \max_{j \in A} c_j \tag{2.68}$$

is submodular and normalized (we take $f(\emptyset) = 0$).

## Proof.

Consider

$$\max_{j \in A} c_j + \max_{j \in B} c_j \geq \max_{j \in A \cup B} c_j + \max_{j \in A \cap B} c_j \tag{2.69}$$

which follows since we have that

$$\max(\max_{j \in A} c_j, \max_{j \in B} c_j) = \max_{j \in A \cup B} c_j \tag{2.70}$$

and

$$\min(\max_{j \in A} c_j, \max_{j \in B} c_j) \geq \max_{j \in A \cap B} c_j \tag{2.71}$$

□

# Max

Given $V$, let $c \in \mathbb{R}^V$ be a given fixed vector (not necessarily non-negative). Then $f : 2^V \to \mathbb{R}$, where

$$f(A) = \max_{j \in A} c_j \tag{2.72}$$

is submodular, where we take $f(\emptyset) \leq \min_j c_j$ (so the function is not normalized).

**Proof.**

The proof is identical to the normalized case. □

# Facility/Plant Location (uncapacitated)

- Let $F = \{1, \dots, f\}$ be a set of possible factory/plant locations for facilities to be built.
- $S = \{1, \dots, s\}$ is a set of sites needing to be serviced (e.g., cities, clients).
- Let $c_{ij}$ be the "benefit" (e.g., $1/c_{ij}$ is the cost) of servicing site $i$ with facility location $j$.
- Let $m_j$ be the benefit (e.g., either $1/m_j$ is the cost or $-m_j$ is the cost) to build a plant at location $j$.
- Each site needs to be serviced by only one plant but no less than one.
- Define $f(A)$ as the "delivery benefit" plus "construction benefit" when the locations $A \subseteq F$ are to be constructed.
- We can define $f(A) = \sum_{j \in A} m_j + \sum_{i \in F} \max_{j \in A} c_{ij}$.
- Goal is to find a set $A$ that maximizes $f(A)$ (the benefit) placing a bound on the number of plants $A$ (e.g., $|A| \leq k$).

# Facility Location

Given $V, E$, let $c \in \mathbb{R}^{V \times E}$ be a given $|V| \times |E|$ matrix. Then

$$f : 2^E \to \mathbb{R}, \text{ where } f(A) = \sum_{i \in V} \max_{j \in A} c_{ij} \tag{2.73}$$

is submodular.

**Proof.**

We can write $f(A)$ as $f(A) = \sum_{i \in V} f_i(A)$ where $f_i(A) = \max_{j \in A} c_{ij}$ is submodular (max of a $i^{\text{th}}$ row vector), so $f$ can be written as a sum of submodular functions. □

Thus, the facility location function (which only adds a modular function to the above) is submodular.

# Log Determinant

- Let $\mathbf{\Sigma}$ be an $n \times n$ positive definite matrix. Let $V = \{1, 2, \ldots, n\} \equiv [n]$ be an index set, and for $A \subseteq V$, let $\mathbf{\Sigma}_A$ be the (square) submatrix of $\mathbf{\Sigma}$ obtained by including only entries in the rows/columns given by $A$.

**Proof.**

Suppose $x \in \mathbf{R}^n$ is multivariate Gaussian, that is

$$x \in p(x) = \frac{1}{\sqrt{|2\pi\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(x-\mu)^T \mathbf{\Sigma}^{-1}(x-\mu)\right) \tag{2.75}$$

...

# Log Determinant

- Let $\mathbf{\Sigma}$ be an $n \times n$ positive definite matrix. Let $V = \{1, 2, \ldots, n\} \equiv [n]$ be an index set, and for $A \subseteq V$, let $\mathbf{\Sigma}_A$ be the (square) submatrix of $\mathbf{\Sigma}$ obtained by including only entries in the rows/columns given by $A$.
- We have that:

$$f(A) = \log \det(\mathbf{\Sigma}_A) \text{ is submodular.} \tag{2.74}$$

**Proof.**

Suppose $x \in \mathbf{R}^n$ is multivariate Gaussian, that is

$$x \in p(x) = \frac{1}{\sqrt{|2\pi\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(x-\mu)^T \mathbf{\Sigma}^{-1}(x-\mu)\right) \tag{2.75}$$

...

# Log Determinant

- Let $\mathbf{\Sigma}$ be an $n \times n$ positive definite matrix. Let $V = \{1, 2, \ldots, n\} \equiv [n]$ be an index set, and for $A \subseteq V$, let $\mathbf{\Sigma}_A$ be the (square) submatrix of $\mathbf{\Sigma}$ obtained by including only entries in the rows/columns given by $A$.
- We have that:

$$f(A) = \log\det(\mathbf{\Sigma}_A) \text{ is submodular.} \tag{2.74}$$

- The submodularity of the log determinant is crucial for determinantal point processes (DPPs) (defined later in the class).

**Proof.**

Suppose $x \in \mathbf{R}^n$ is multivariate Gaussian, that is

$$x \in p(x) = \frac{1}{\sqrt{|2\pi\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(x-\mu)^T \mathbf{\Sigma}^{-1}(x-\mu)\right) \tag{2.75}$$

...

# Log Determinant

**...cont.**

Then the (differential) entropy of the r.v. $X$ is given by

$$h(X) = \log \sqrt{|2\pi e \mathbf{\Sigma}|} = \log \sqrt{(2\pi e)^n |\mathbf{\Sigma}|} \tag{2.76}$$

and in particular, for a variable subset $A$,

$$f(A) = h(X_A) = \log \sqrt{(2\pi e)^{|A|} |\mathbf{\Sigma}_A|} \tag{2.77}$$

Entropy is submodular (conditioning reduces entropy), and moreover

$$f(A) = h(X_A) = m(A) + \frac{1}{2} \log |\mathbf{\Sigma}_A| \tag{2.78}$$

where $m(A)$ is a modular function. □

Note: still submodular in the semi-definite case as well.

# Concave over non-negative modular

Let $m \in \mathbb{R}_+^E$ be a modular function, and $g$ a concave function over $\mathbb{R}$. Define $f : 2^E \to \mathbb{R}$ as

$$f(A) = g(m(A)) \tag{2.79}$$

then $f$ is submodular.

**Proof.**

Given $A \subseteq B \subseteq E \setminus v$, we have $0 \leq a = m(A) \leq b = m(B)$, and $0 \leq c = m(v)$. For $g$ concave, we have $g(a+c) - g(a) \geq g(b+c) - g(b)$, and thus

$$g(m(A) + m(v)) - g(m(A)) \geq g(m(B) + m(v)) - g(m(B)) \tag{2.80}$$

$\square$

A form of converse is true as well.