Submodularity and its application in Social Network Analysis

Attila Kiss

Department of Information Systems

Eötvös Loránd University, Budapest, Hungary

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013) and with the support of the Hungarian and Vietnamese TET (grant agreement no. TET 10-1-2011-0645).

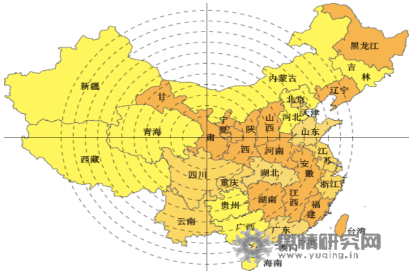
Vietnam, 2014

Outline

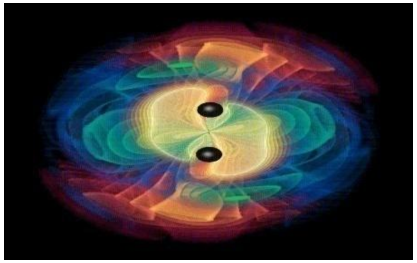
- Social networks
- Information diffusion and social effect maximization
- Submodular functions and their applications

Social Networks

SN bridges our daily life and the **virtual** web space!

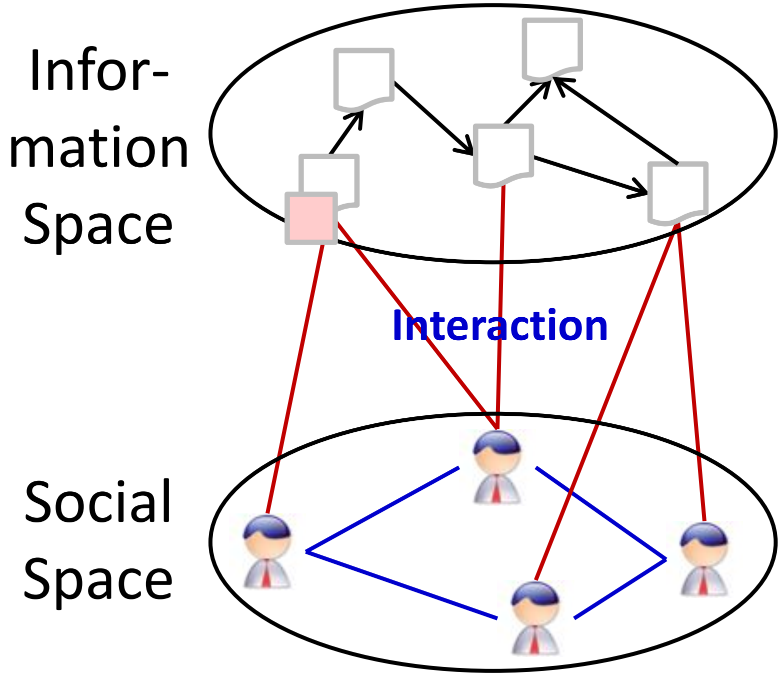


Opinion Mining



Innovation diffusion

Business Intelligence



Information→ user

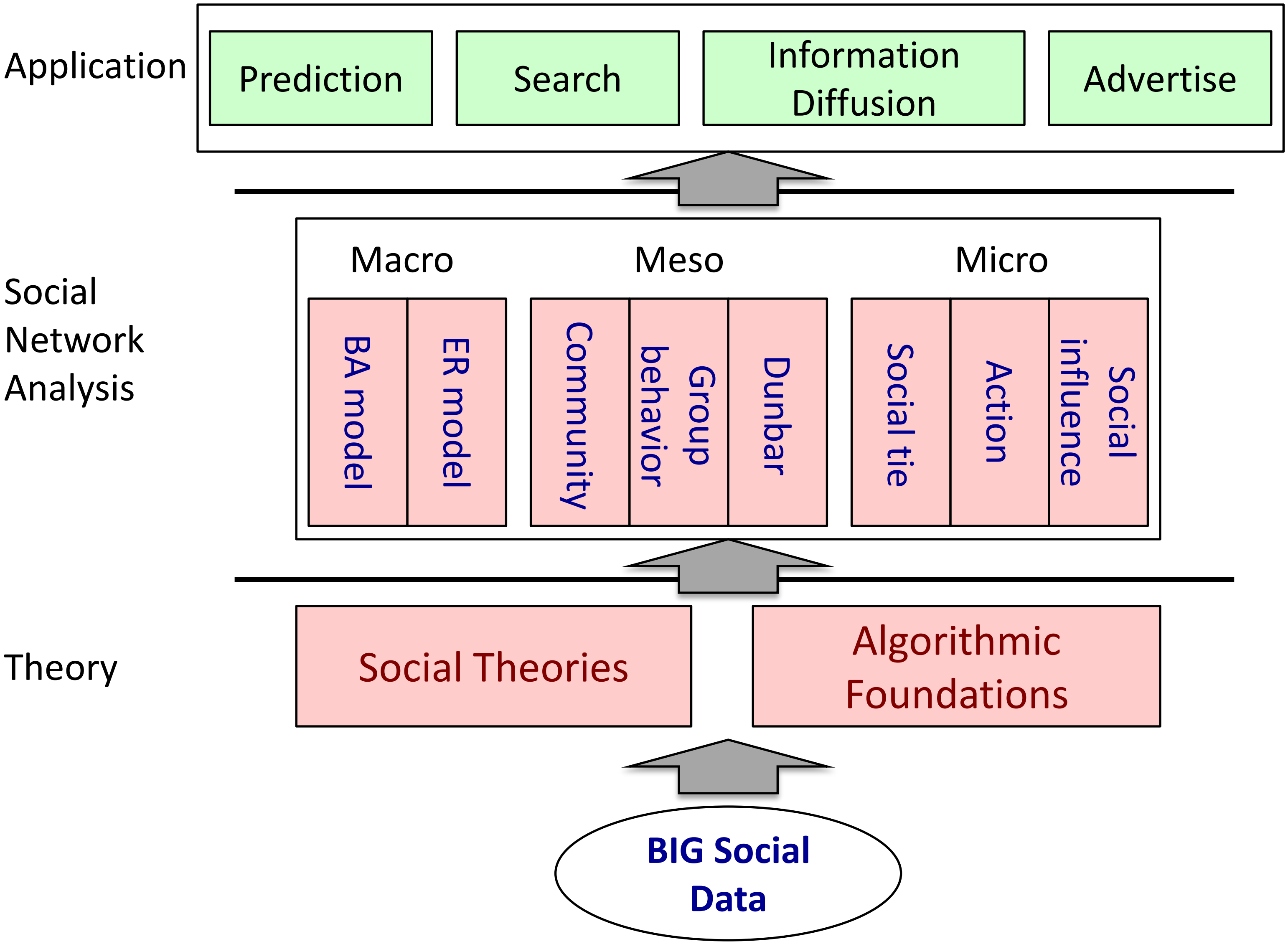
Interaction mechanism

facebook

twitter

Overview of Core Research in Social Networks

Core Research in Social Network



Application

Prediction

Search

Information Diffusion

Advertise

Social Network Analysis

Macro

BA model

ER model

Meso

Community

Group behavior

Dunbar

Micro

Social tie

Action

Social influence

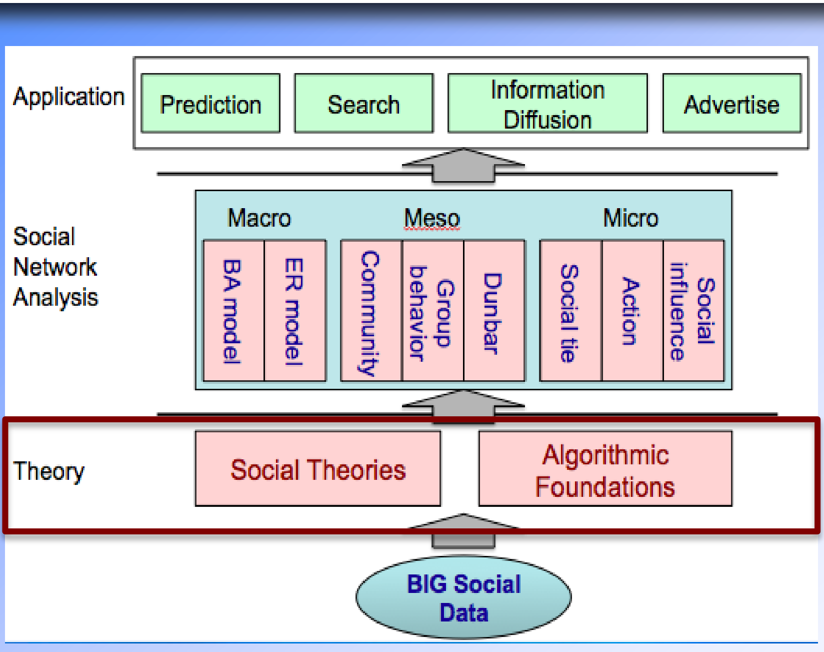
Theory

Social Theories

Algorithmic Foundations

BIG Social Data

Computational Foundations for Social Networks

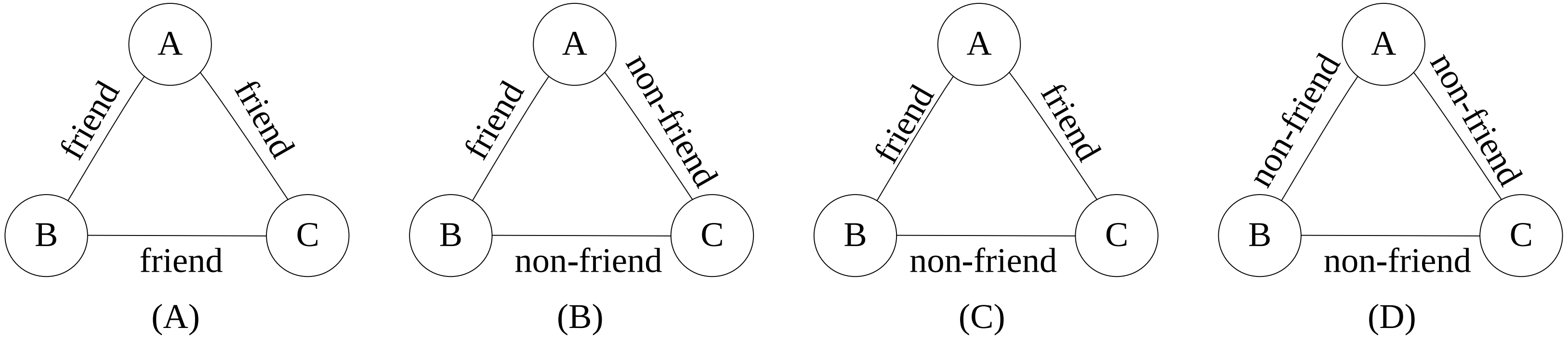


Computational Foundations

- Social Theories
  - Social balance
  - Social status
  - Structural holes
  - Two-step flow
- Algorithmic Foundations
  - Network flow
  - K-densest subgraph
  - Set cover

Social Theories—Social Balance

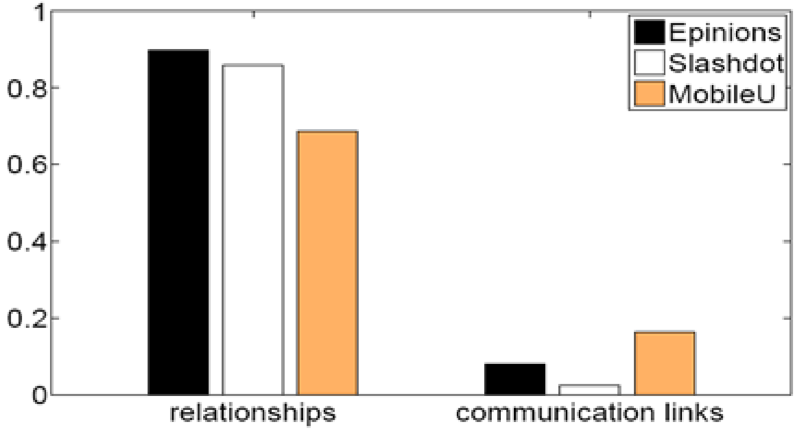
Your friend's friend is your friend, and your enemy's enemy is also your friend.



Examples on Epinions, Slashdot, and MobileU

(1) The underlying networks are unbalanced;

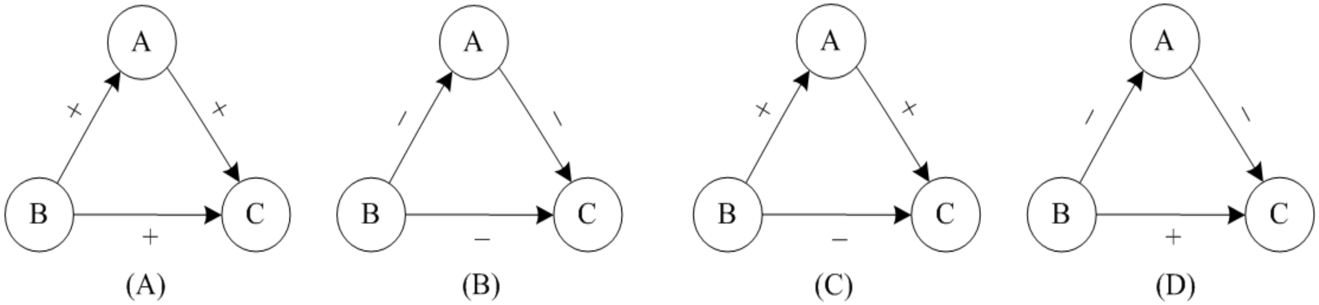
(2) While the friendship networks are balanced.



Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring Social Ties across Heterogeneous Networks. In WSDM'2012. pp. 743-752.

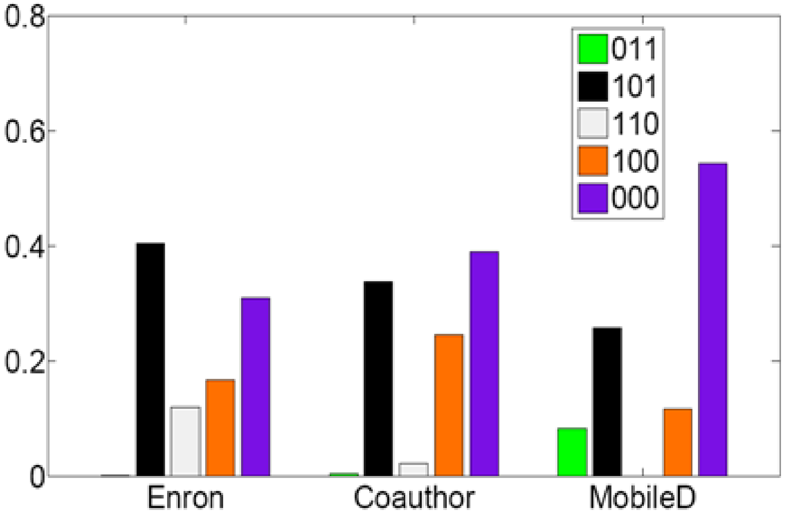
Social Theories—Social status

Your boss's boss is also your boss…



Observations: 99% of triads in the networks satisfy the social status theory

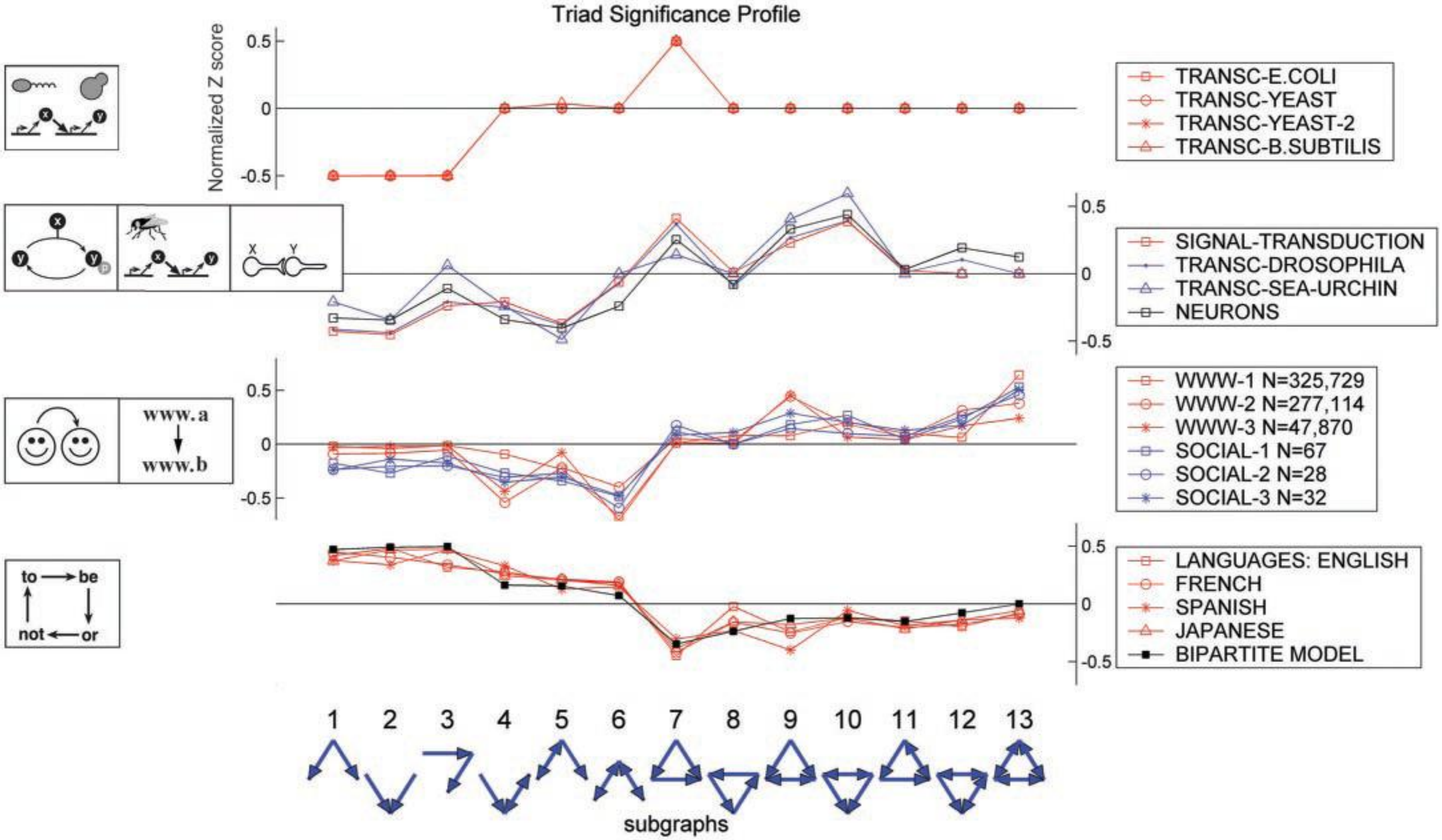
Examples: Enron, Coauthor, MobileD



Note: Given a triad (A,B,C), let us use 1 to denote the advisor-advisee relationship and 0 colleague relationship. Thus the number 011 to denote A and B are colleagues, B is C's advisor and A is C's advisor.

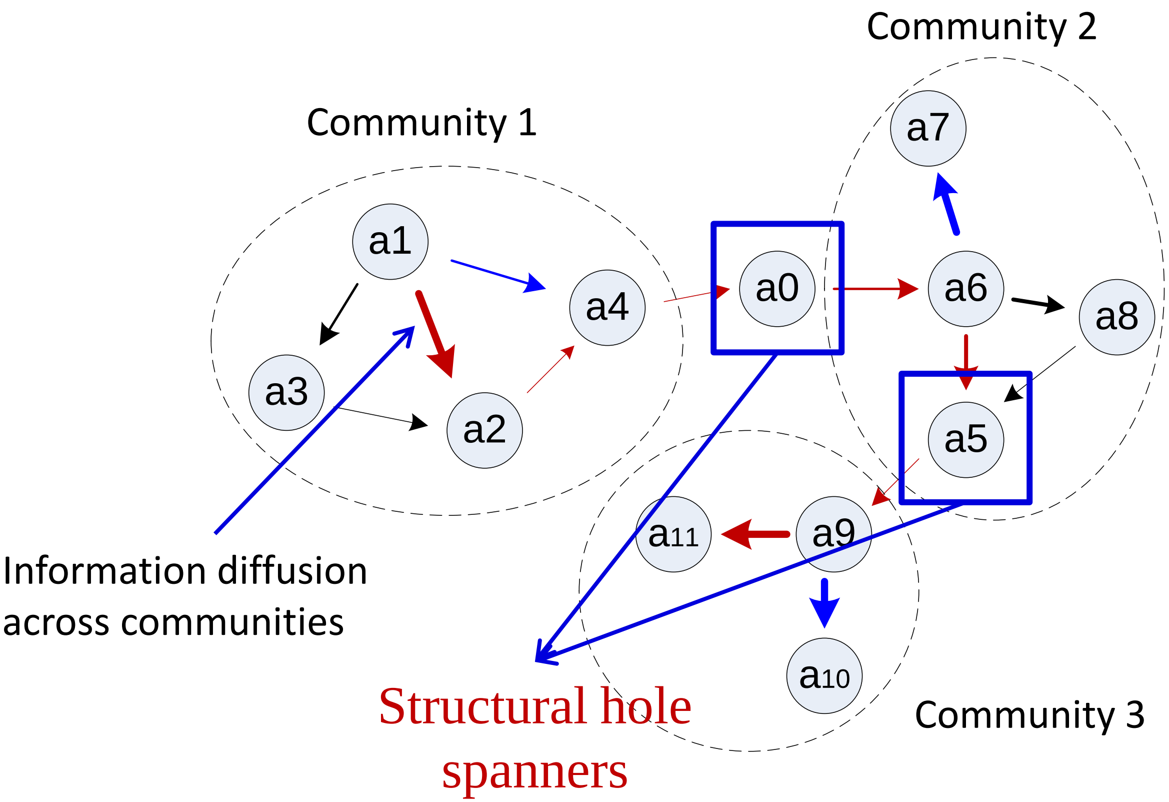
Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring Social Ties across Heterogeneous Networks. In WSDM'2012. pp. 743-752.

Triadic Closure



R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. Science, 2004

Social Theories—Structural holes



twitter

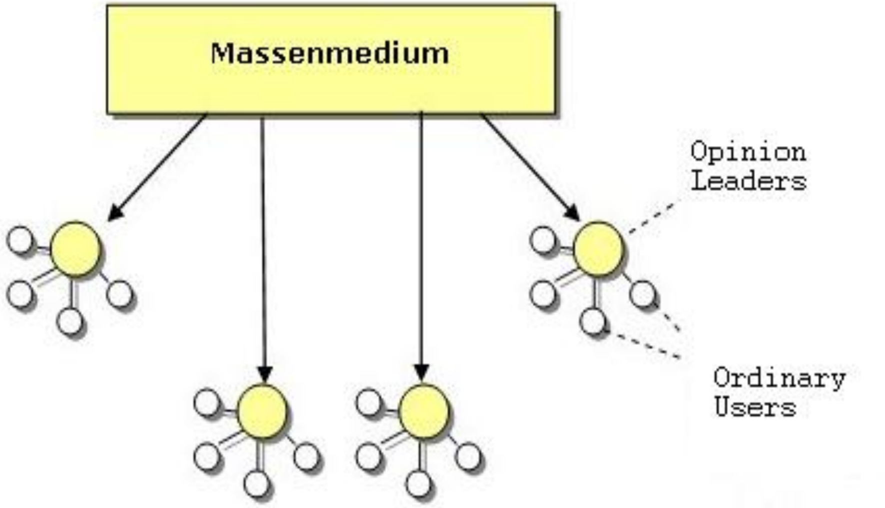
1% twitter users control **25%** retweeting behaviors between communities.

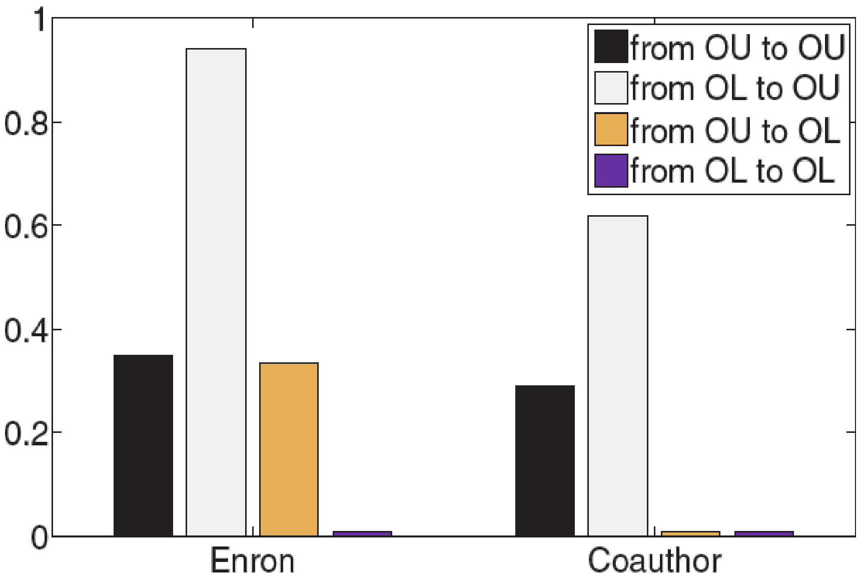
Structural hole users control the information flow between different communities (Burt, 92; Podolny, 97; Ahuja, 00; Kleinberg, 08; Lou & Tang, 13)

T. Lou and J. Tang. Mining Structural Hole Spanners Through Information Diffusion in Social Networks. In **WWW'13**. pp. 837-848.

Social Theories—Two-step-flow

Lazarsfeld *et al* suggested that:
"ideas often flow from radio and print to the opinion leaders and from them to the less active sections of the population."





Estimate OL and OU by PageRank
OL : Opinion leader;
OU : Ordinary user.

Observations: Opinion leaders are more likely (+71%-84% higher than chance) to spread information to ordinary users.

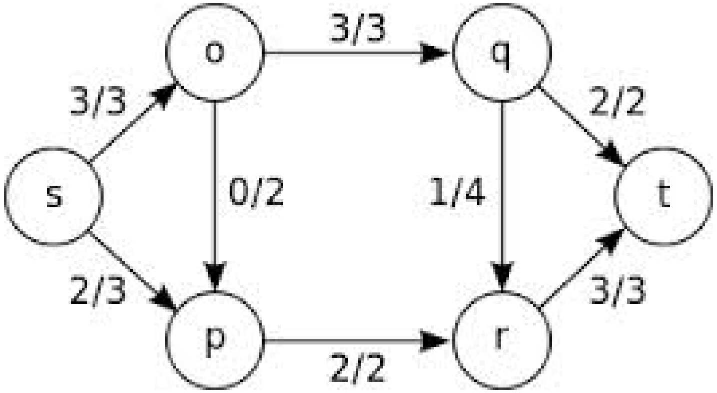
Lazarsfeld et al. (1944). The people's choice: How the voter makes up his mind in a presidential campaign.

Computational Foundations

- Social Theories
  - Social balance
  - Social status
  - Structural holes
  - Two-step flow
- Algorithmic Foundations
  - Network flow
  - K-densest subgraph
  - Set cover

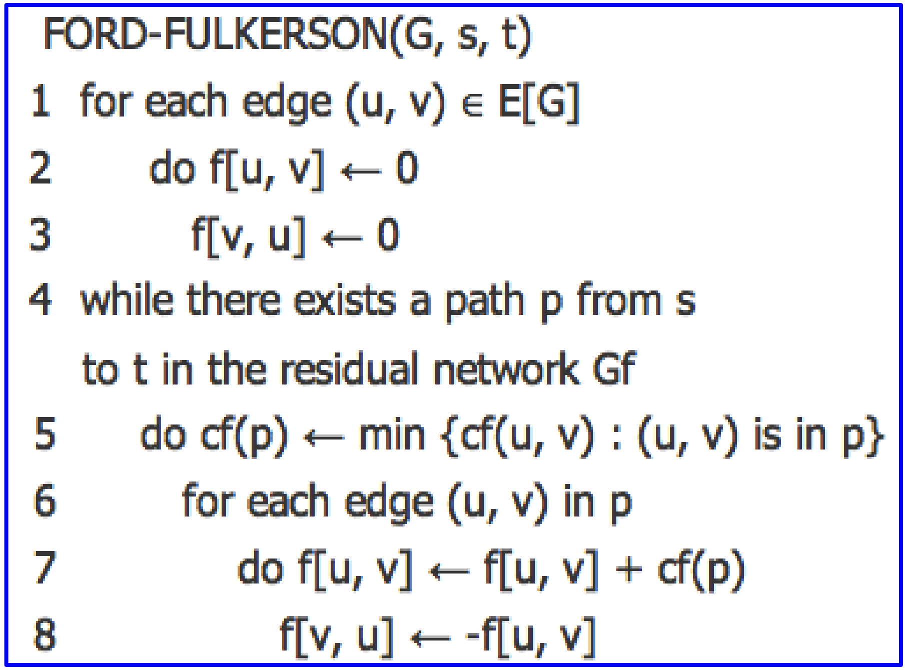
Algorithm — Network Flow

- Classical problems:
  - Maximum flow / minimum cut
    - Ford-Fulkerson algorithm
    - Dinic algorithm
  - Minimum cut between multiple sets of vertices
    - NP hard when there are more than 2 sets
  - Minimum cost flow;
  - Circulation problem;
  - …



Algorithm — Network Flow (cont.)

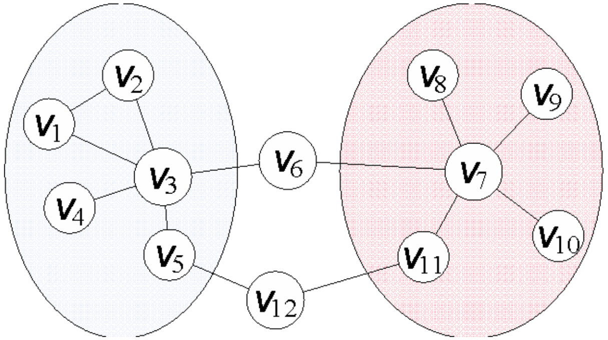
- Ford-Fulkerson
    - As long as there is an augmenting path, send the minimum of the residual capacities on the path.
    - A maximum flow is obtained when no augmenting paths left.
    - Time complexity: O(VE^2)



```
FORD-FULKERSON(G, s, t)
for each edge (u, v) ∊ E[G]
    do f[u, v] ← 0
       f[v, u] ← 0
while there exists a path p from s
to t in the residual network Gf
    do cf(p) ← min {cf(u, v) : (u, v) is in p}
       for each edge (u, v) in p
           do f[u, v] ← f[u, v] + cf(p)
              f[v, u] ← -f[u, v]
```

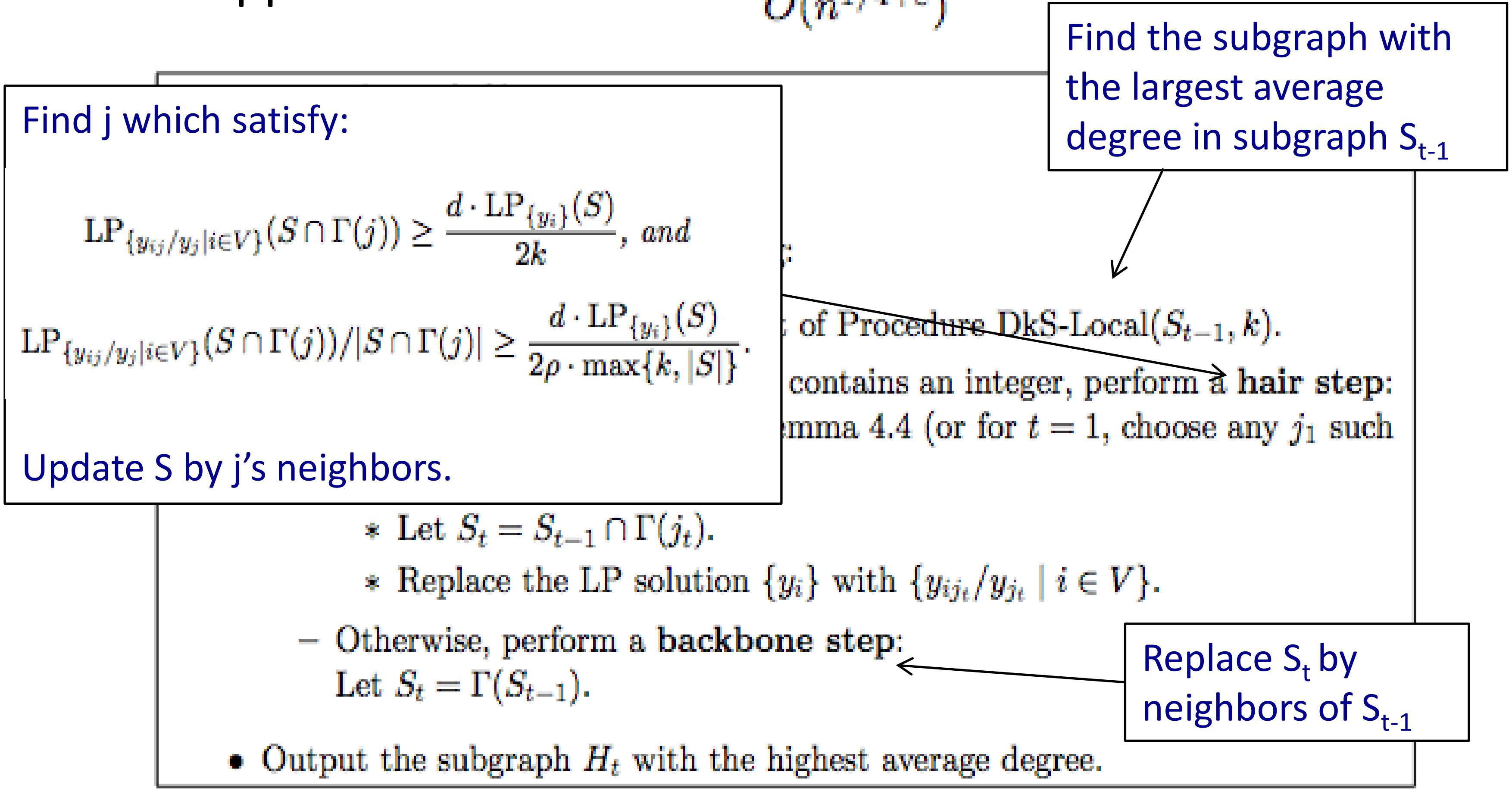
Algorithm — K-densest subgraph

- NP Problem
  - Find the maximum density subgraph on exactly k vertices.
  - Reduced from the clique problem
- Application
  - Reduce the structural hole spanner detection problem to proof its NP hardness.
  - To find a subset of nodes, such that without them, the connection between communities would be minimized.



Algorithm — K-densest subgraph (cont.)

- A linear programming based solution
  - Approximation ratio: $O(n^{1/4+\varepsilon})$



Find j which satisfy:

$$\mathrm{LP}_{\{y_{ij}/y_j \mid i \in V\}}(S \cap \Gamma(j)) \geq \frac{d \cdot \mathrm{LP}_{\{y_i\}}(S)}{2k}, \text{ and}$$

$$\mathrm{LP}_{\{y_{ij}/y_j \mid i \in V\}}(S \cap \Gamma(j))/|S \cap \Gamma(j)| \geq \frac{d \cdot \mathrm{LP}_{\{y_i\}}(S)}{2\rho \cdot \max\{k, |S|\}}.$$

Update S by j's neighbors.

Find the subgraph with the largest average degree in subgraph S_{t-1}

:

of Procedure DkS-Local(S_{t-1}, k).

contains an integer, perform a **hair step**:

emma 4.4 (or for $t = 1$, choose any j_1 such

- ∗ Let $S_t = S_{t-1} \cap \Gamma(j_t)$.
- ∗ Replace the LP solution $\{y_i\}$ with $\{y_{ij_t}/y_{j_t} \mid i \in V\}$.

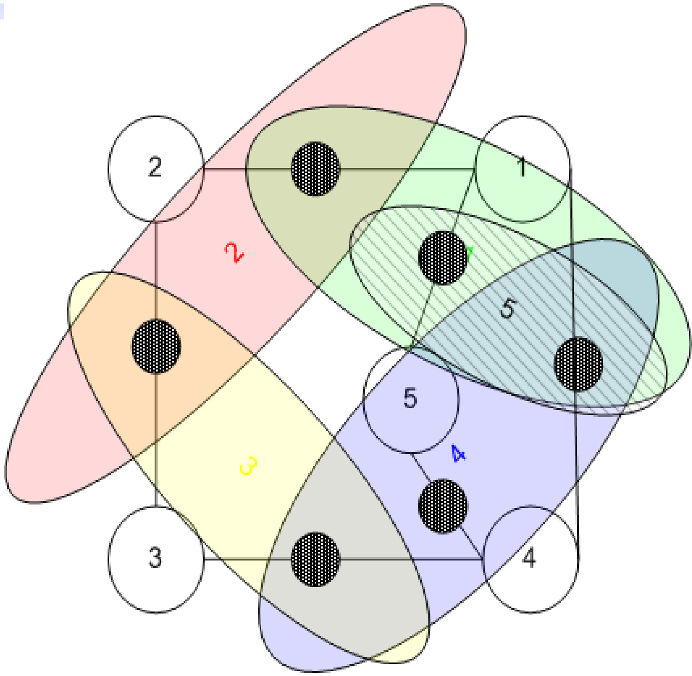
– Otherwise, perform a **backbone step**:
Let $S_t = \Gamma(S_{t-1})$.

- Output the subgraph H_t with the highest average degree.

Replace S_t by neighbors of S_{t-1}

Algorithm — Set Cover

- Another NP problem
  - Given a set of elements (universe) and a set *S* of *n* sets whose union equals the universe;
  - Find the smallest subset of *S* that contains all elements in the universe;
  - The decision version is NP-complete.

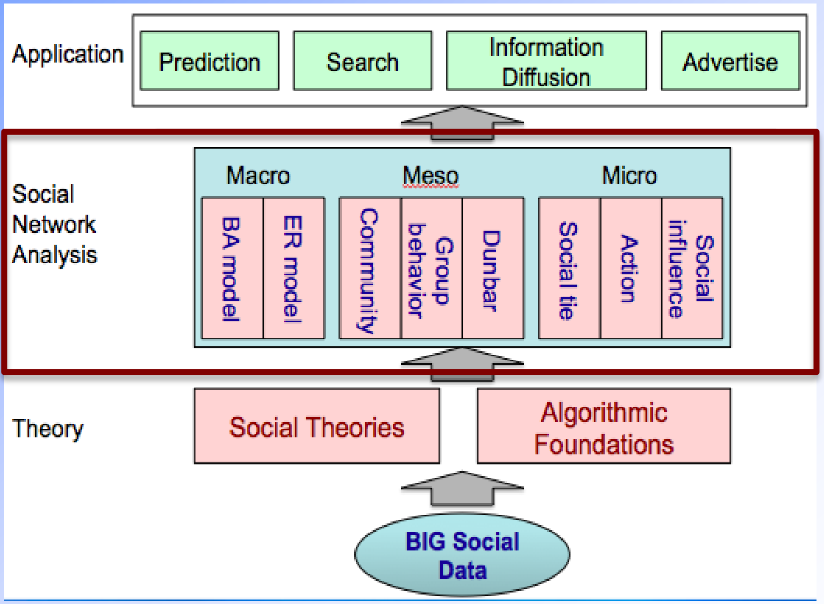


- Greedy
  - Choose the set containing the most uncovered elements;
  - Approximation ratio: *H*(size(*S*)), where *H*(*n*) is the n-th harmonic number.

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{k=1}^{n} \frac{1}{k}.$$

Social Network Analysis

- Macro Level
- Meso Level
- Micro Level



Erdős–Rényi Model

In the *G*(*n*, *p*) model, each edge is included in the graph with probability p independent from every other edge.

Each random graph has the probability

$$p^M(1-p)^{\binom{n}{2}-M}.$$

- Properties

(1) Degree distribution-Poisson

$$p(k) = \frac{<k>^k}{k!} e^{-<k>}$$

(2) Clustering coefficient ⟶ Small

$$p$$

(3) Average shortest path

$$L \sim \frac{\ln N}{\ln <k>}$$

Problem: In real social network, neighbors tend to be connected with each other, thus the clustering coefficient should not be too small.

Erdős, P.; Rényi, A. (1959), “On Random Graphs.”.

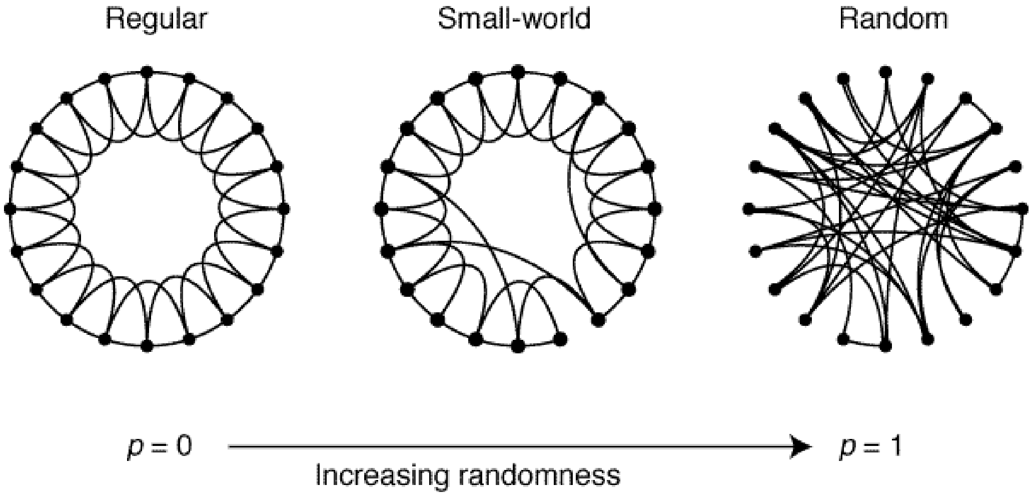
Small-World Model

Mechanism

1. Start from a regular wired ring, where each node is connected with its K-nearest neighbors

2. With probability p rewire each edge.

Problem: In real social network, degree distribution is power law.



- Properties

(1) Degree distribution

$$p(k) = \begin{cases} 0, k < K \longrightarrow \\ \frac{< d >}{(k - K)!} e^{-<d>}, k \geq K \end{cases} \qquad < d >= Kp$$

Not power law

(2) Clustering coefficient

$$C = \frac{3(K - 2)}{4(K - 1) + 4Kp(p + 2)}$$

(3) Average shortest path

$$L = \frac{\ln NKp}{K^2 p}$$

Watts, D. J.; Strogatz, S. H. (1998). "Collective dynamics of 'small-world' networks". Nature 393 (6684): 440–442.

Barabási-Albert Model

Idea

- Growth
- Preferential attachment (rich-get-richer, the Matthew Effect)

Mechanism

1. Start from a small connected graph with m_0 nodes
2. At each time step, add one new node with m ($m \leq m_0$) new edges; the probability that the new node is connected to node i is p_i = $k_i / \sum_j k_j$.

- Degree distribution

$$p(k) = 2m^2 k^{-3}$$

Scale-free

- Clustering coefficient

$$C \sim \frac{(\ln t)^2}{t}$$

- Average longest shortest path

$$L \sim \frac{\ln N}{\ln \ln N}$$

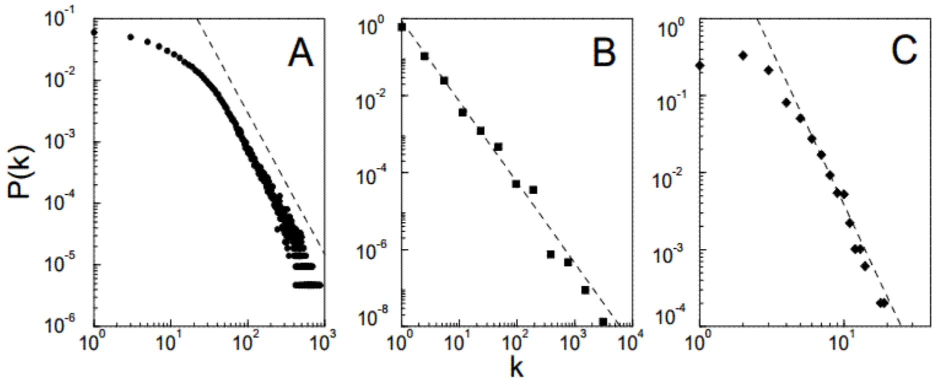
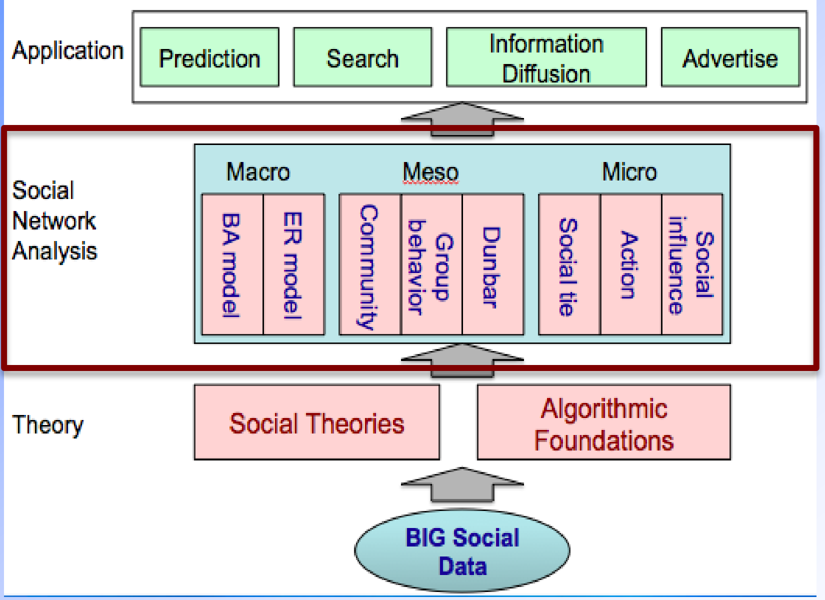


FIG. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212,250$ vertices and average connectivity $\langle k \rangle = 28.78$; (**B**) World wide web, $N = 325,729$, $\langle k \rangle = 5.46$ (6); (**C**) Powergrid data, $N = 4,941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (**A**) $\gamma_{actor} = 2.3$, (**B**) $\gamma_{www} = 2.1$ and (**C**) $\gamma_{power} = 4$.

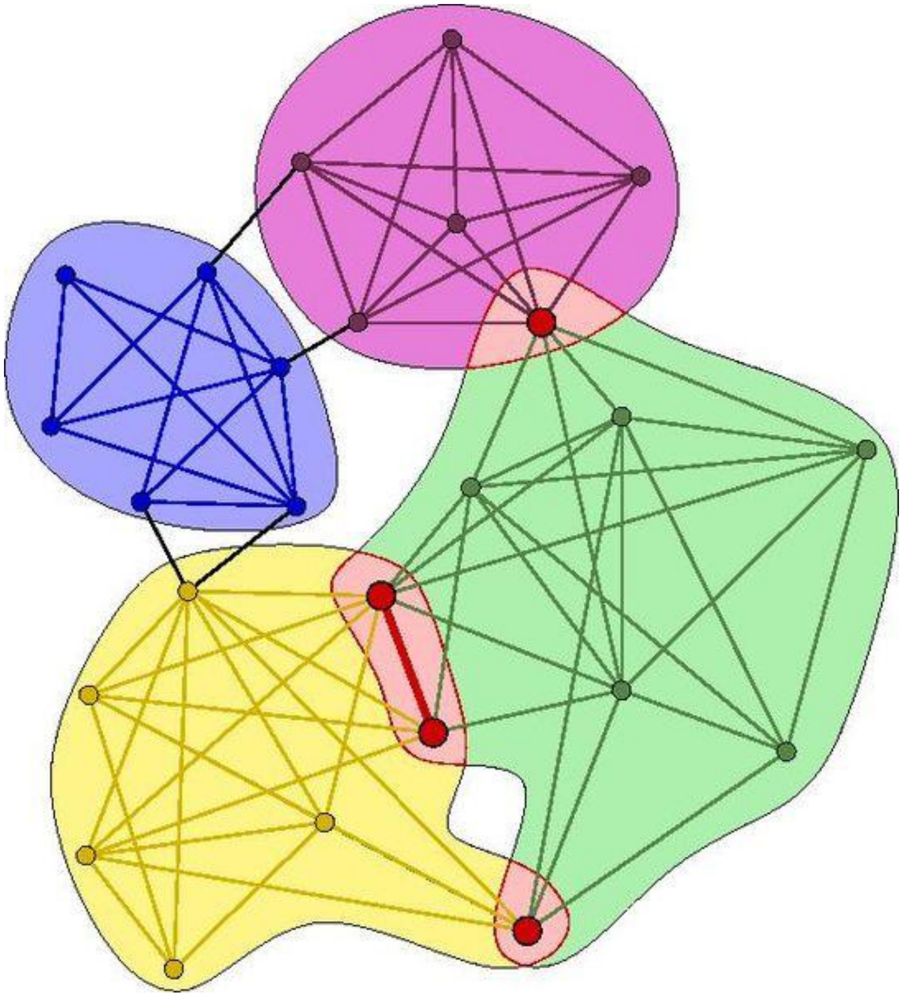
Barabasi and Albert(1999). Emergence of scaling n complex networks.

Social Network Analysis

- Macro Level
- Meso Level
- Micro Level



Community Detection



Node-Centric Community

Each node in a group satisfies certain properties

Group-Centric Community

Consider the connections within a group as a whole. The group has to satisfy certain properties without zooming into node-level

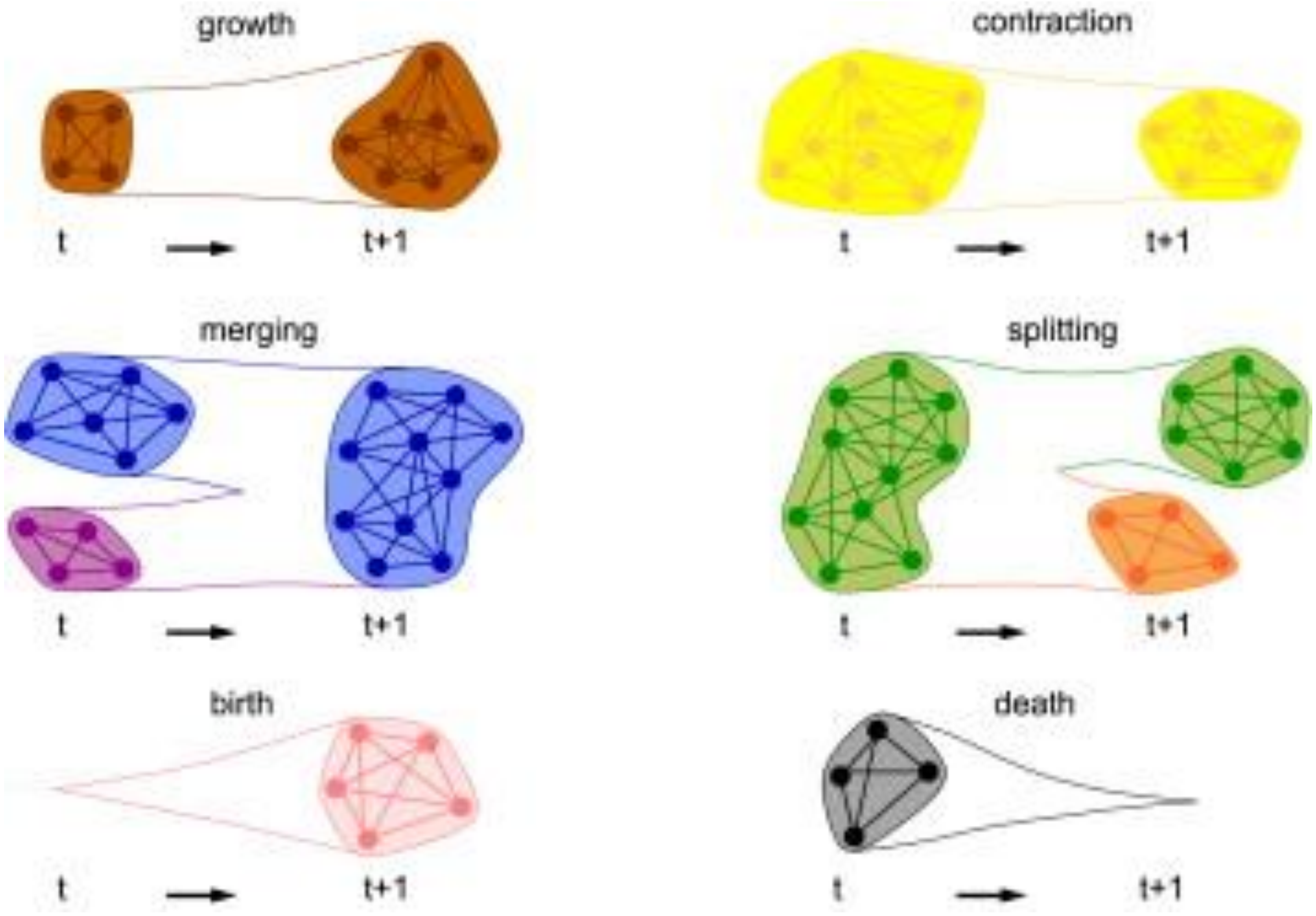
Network-Centric Community

Partition the whole network into several disjoint sets

Hierarchy-Centric Community

Construct a hierarchical structure of communities

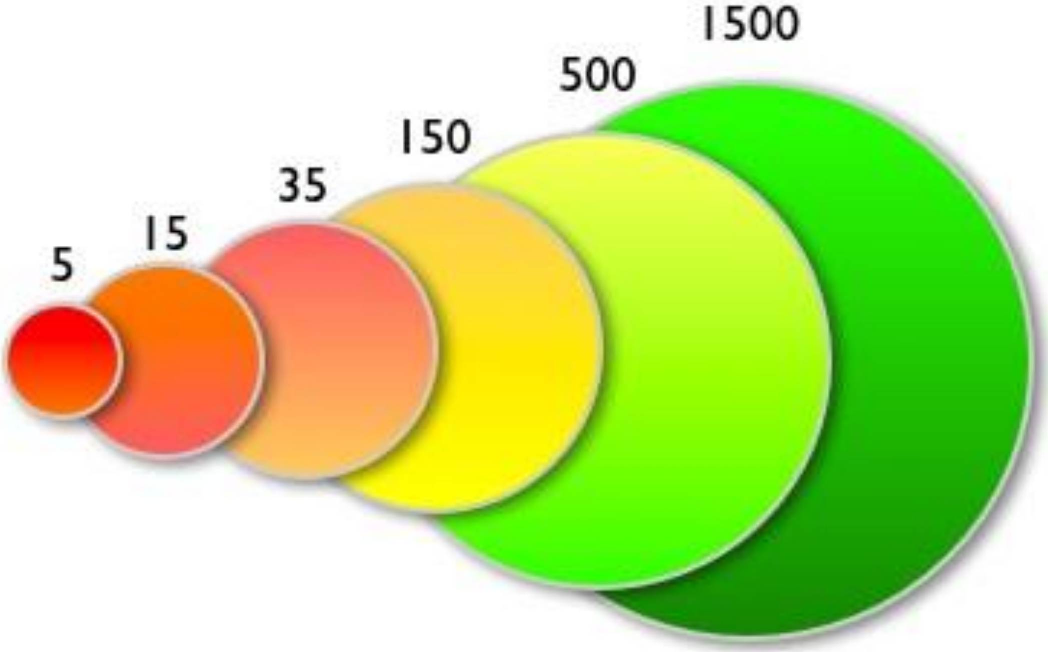
Community Evolution



Dunbar Number

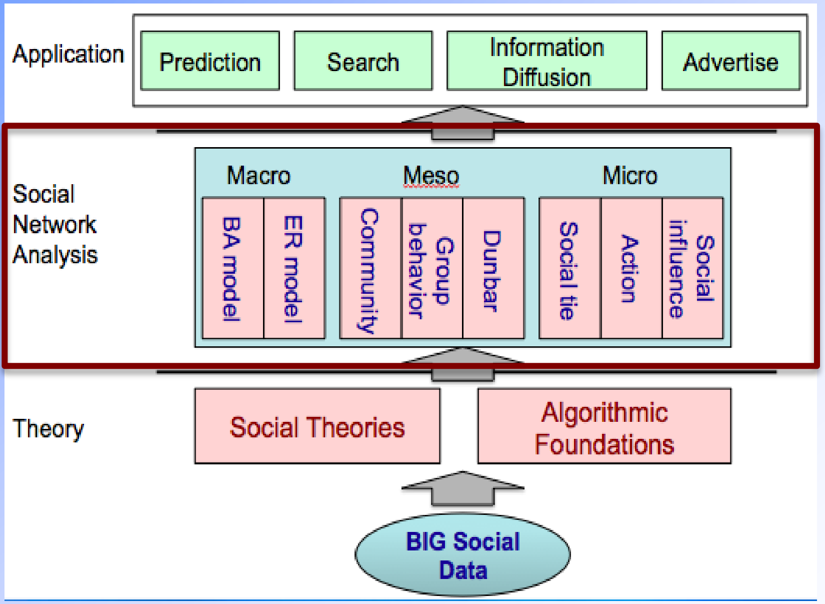
- Dunbar number:150. Dunbar's number is a suggested cognitive limit to the number of people with whom one can maintain stable social relationships

—Robin Dunbar, 2000



Social Network Analysis

- Macro Level
- Meso Level
- Micro Level



Social Action

- …the object is to interpret the meaning of social action and thereby give a causal explanation of the way in which the action proceeds and the effects which it produces...

— Social Action Theory, by Max Weber, 1922

THEORY AND DECISION LIBRARY · VOLUME 43

SOCIAL ACTION

Edited by Gottfried Seebass and Raimo Tuomela

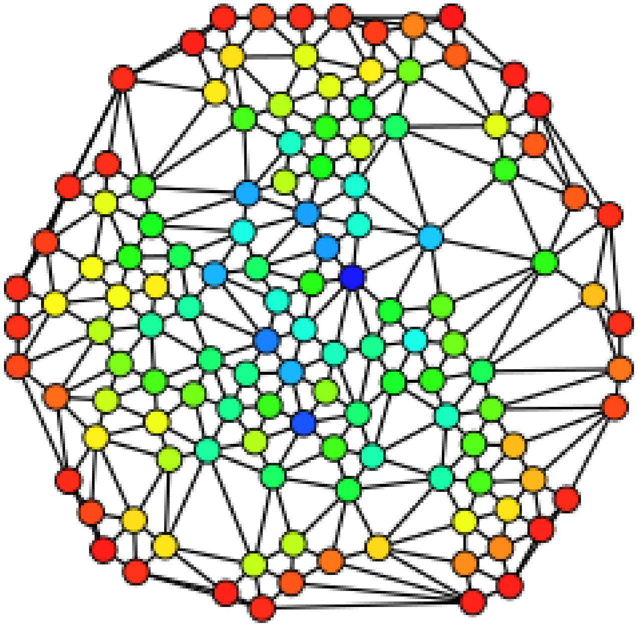
D. REIDEL PUBLISHING COMPANY
DORDRECHT / BOSTON / LANCASTER

Social Action — User Characterization

- Betweenness
  - A centrality measure of a vertex within a graph
  - $$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

#shortest paths pass through v

#shortest paths from s to t



Hue (from red=min to blue=max) shows the node betweenness.

Social Action — User Characterization (cont.)

- Clustering Coefficient
  - A measure of degree to which nodes in a graph tend to cluster together.
  - Global clustering coefficient
    - $$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$
    - A triangle consists of three closed triplets, and a closed triplet consists of three nodes connected to each other.
  - Local clustering coefficient

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

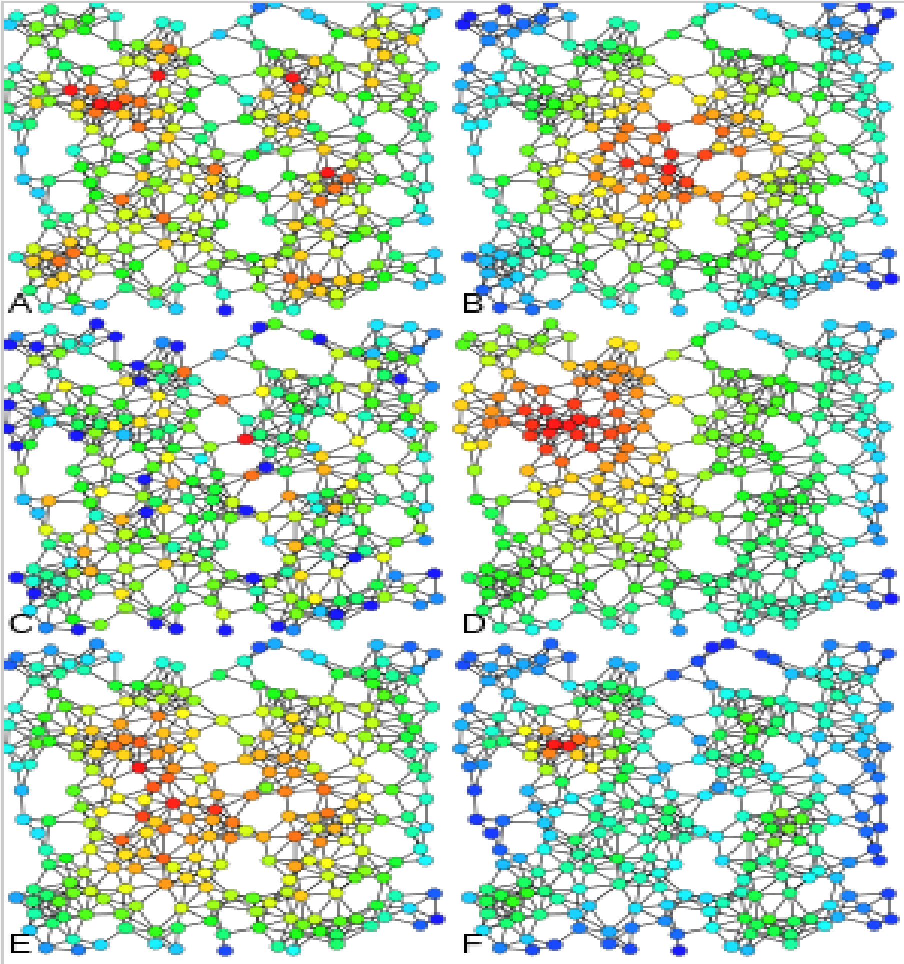
Social Action — User Characterization (cont.)

- Degree: the number of one vertex's neighbors.
- Closeness: the shortest path between one vertex and another vertex.

$$C_C(v) = \sum_{t \in V \setminus v} 2^{-d_G(v,t)}$$

Social Action — User Characterization (cont.)

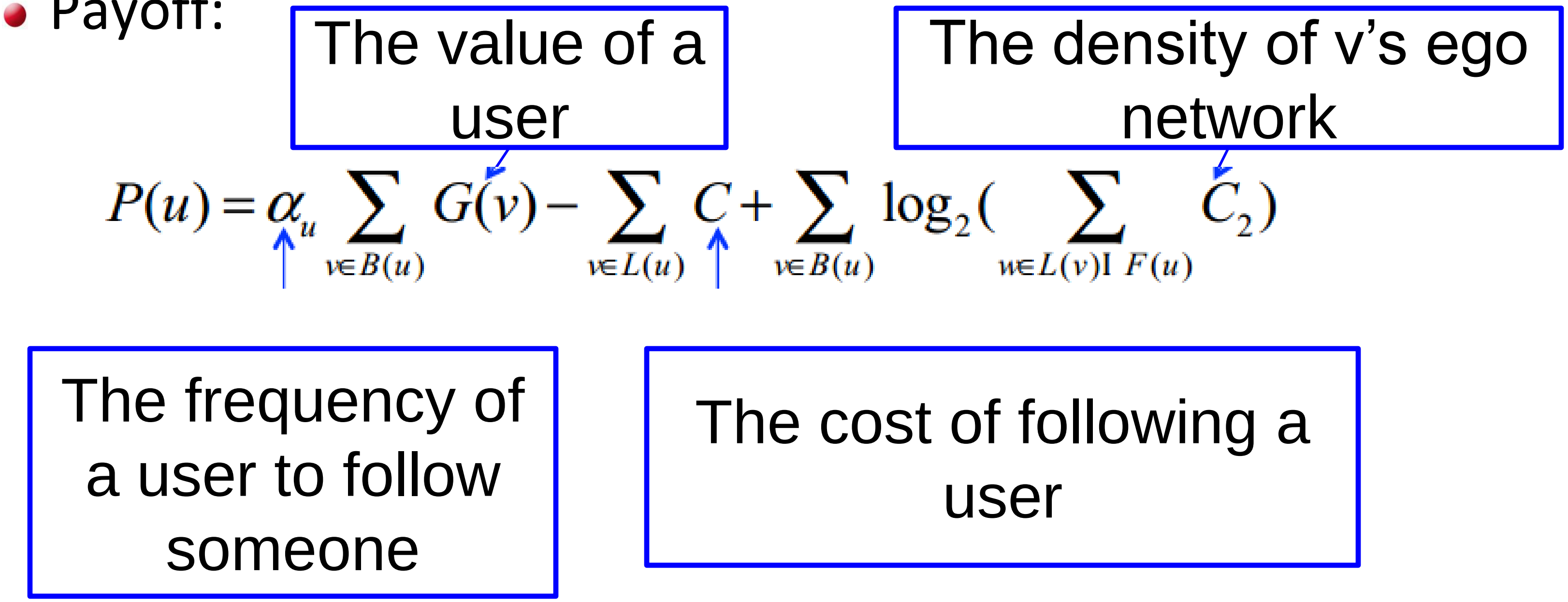
- Centrality



Examples of A) Degree centrality, B) Closeness centrality, C) Betweenness centrality, D) Eigenvector centrality, E) Katz centrality and F) Alpha centrality of the same graph.

Social Action — Game Theory

- Example: a game theory model.
  - Strategy: whether to follow a user or not;
  - Payoff:



The value of a user

The density of v’s ego network

$$P(u) = \alpha_u \sum_{v \in B(u)} G(v) - \sum_{v \in L(u)} C + \sum_{v \in B(u)} \log_2 \Big(\sum_{w \in L(v) \cap F(u)} C_2 \Big)$$

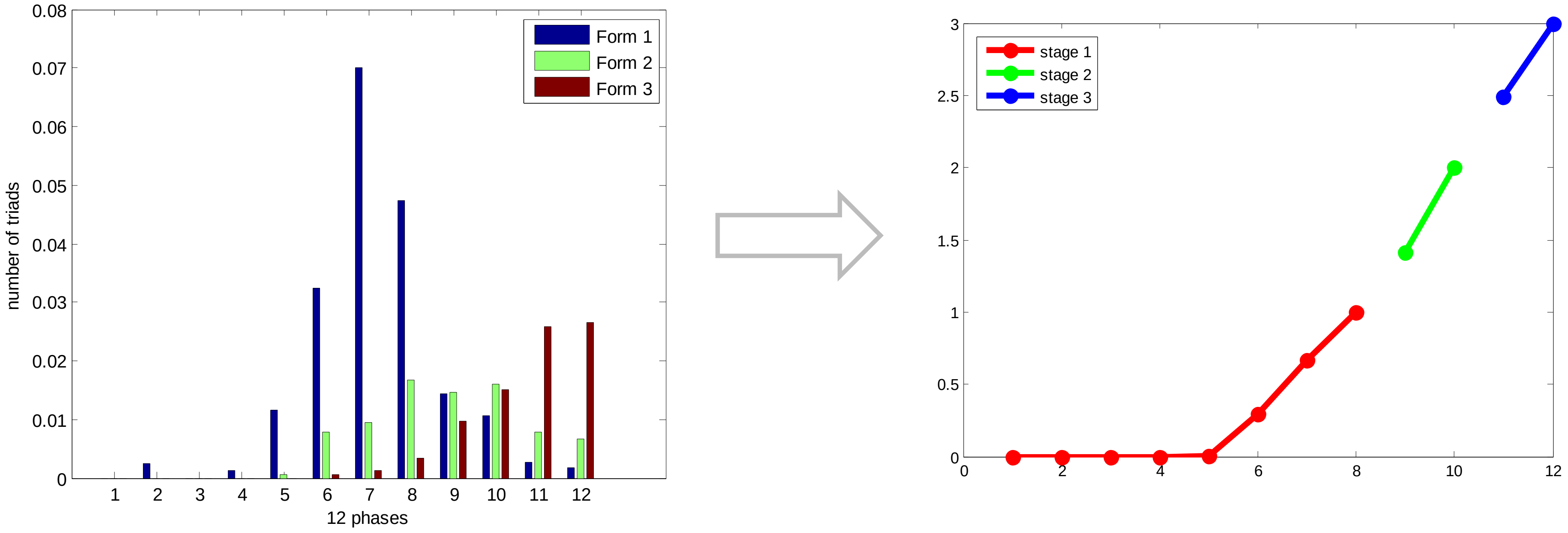
The frequency of a user to follow someone

The cost of following a user

  - The model has a pure strategy Nash Equilibrium

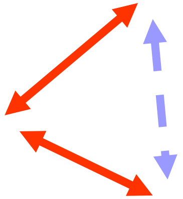
Social Action — Game Theory (cont.)

- Results: three stage life cycle
  - Stage 1: getting into a community
  - Stage 2: becoming an elite
  - Stage 3: bridging different communities (structural hole spanners)

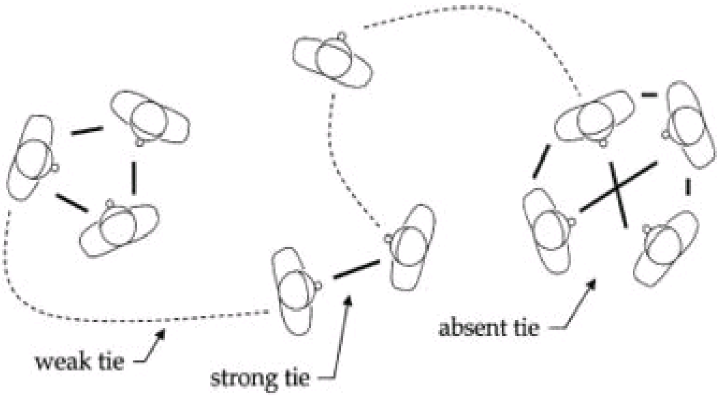


Strong/Weak Ties

- Strong ties
  - Frequent communication, but ties are redundant due to high clustering
- Weak ties
  - Reach far across network, but communication is infrequent…



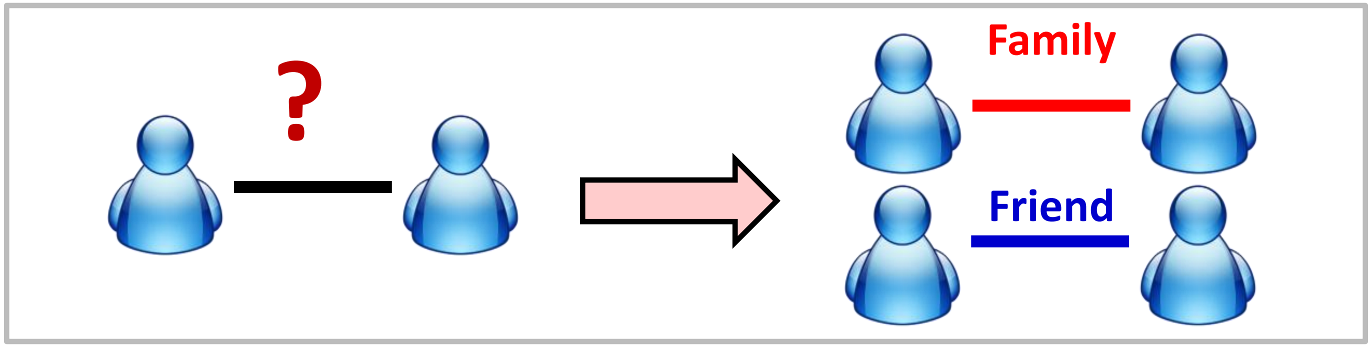
“forbidden triad”：
strong ties are likely to “close”



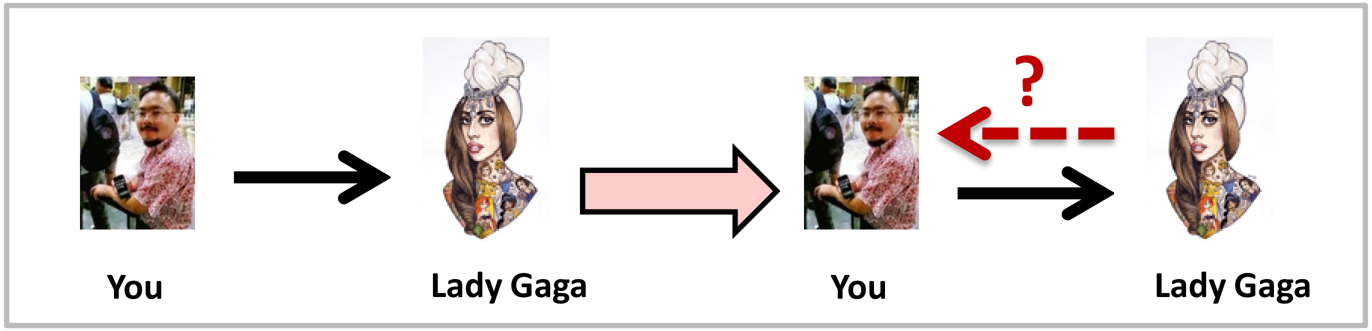
Weak ties act as local bridge

Social Ties

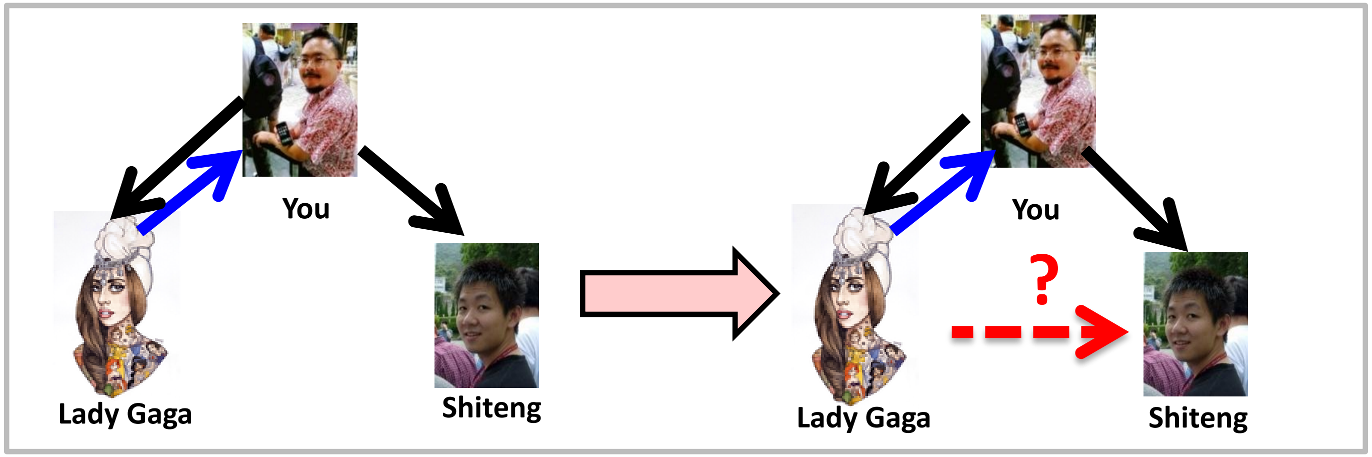
Inferring social ties



Reciprocity



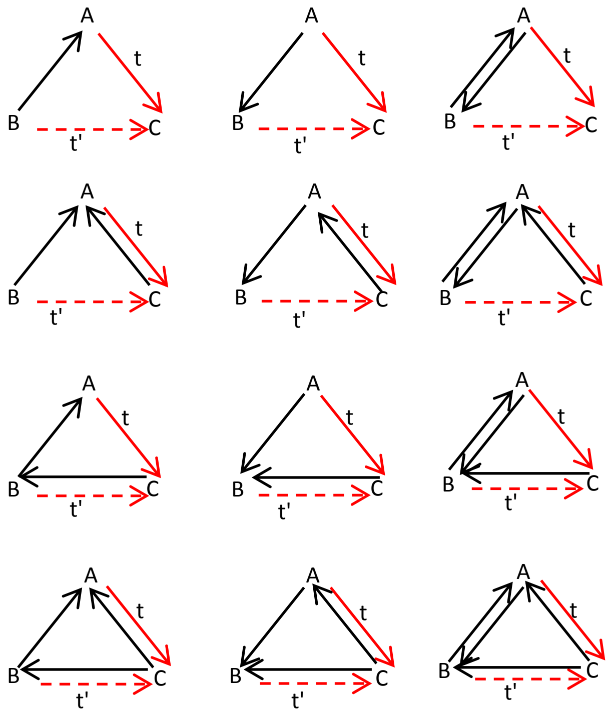
Triadic Closure



KDD 2010, PKDD 2011 (*Best Paper Runnerup*), WSDM 2012, ACM TKDD

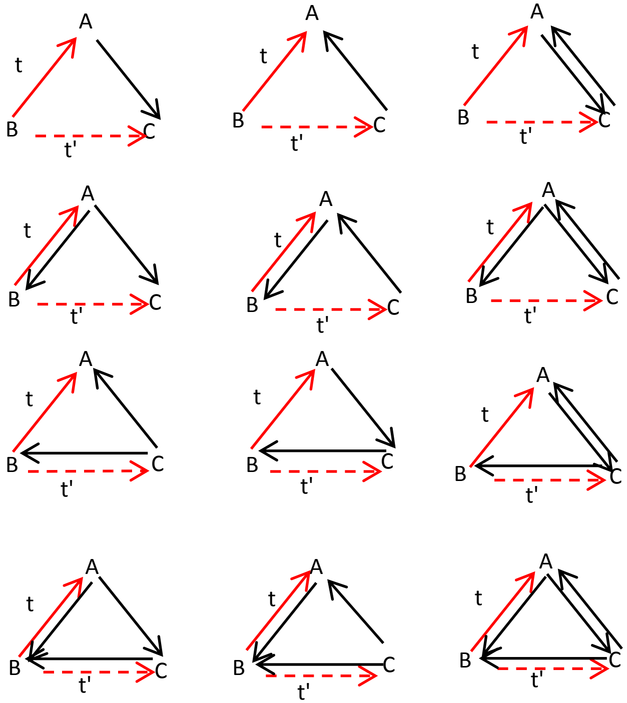
Triadic Closure

Follower diffusion



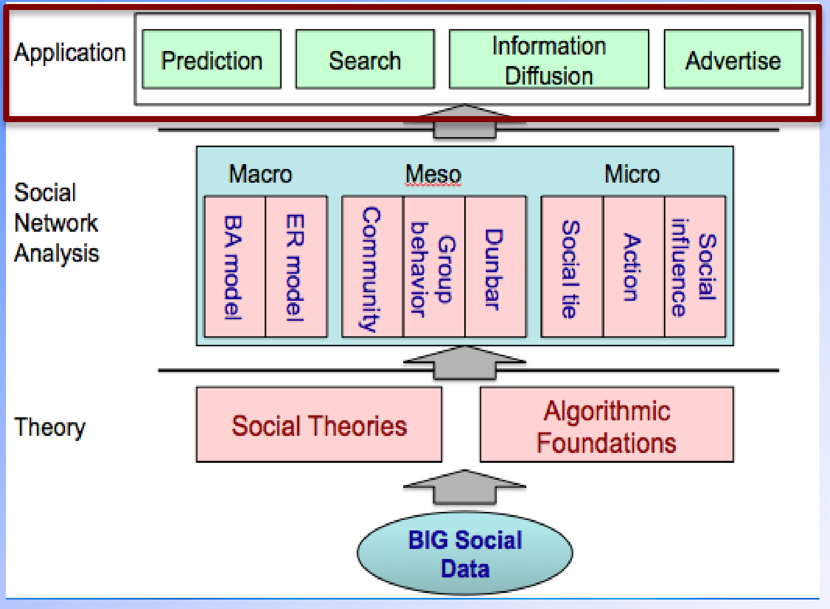
12 triads

Followee diffusion



12 triads

Information Diffusion



Application

Prediction

Search

Information Diffusion

Advertise

Social Network Analysis

Macro

Meso

Micro

BA model

ER model

Community

Group behavior

Dunbar

Social tie

Action

Social influence

Theory

Social Theories

Algorithmic Foundations

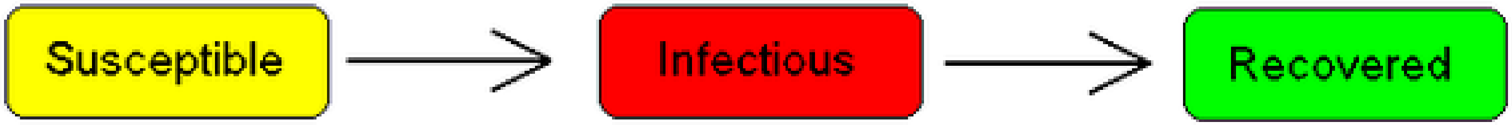
BIG Social Data

Disease-Propagation Models

- Classical disease-propagation models in epidemiology are based upon the cycle of disease in a host.
  - Susceptible
  - Infected
  - Recovered
  - …
- The transition rates from one cycle to another are expressed as derivatives.
- Classical models:
  - SIR
  - SIS
  - SIRS
  - …

SIR Model

- Created by Kermack and McKendrick in 1927.
- Considers three cycles of disease in a host:



Susceptible → Infectious → Recovered

- Transition rates:

$$\frac{dS}{dt} = -\beta S(t)I(t)$$
$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t)$$
$$\frac{dR}{dt} = \gamma I(t)$$

S(t) : #susceptible people at time t;

I(t) : #infected people at time t;

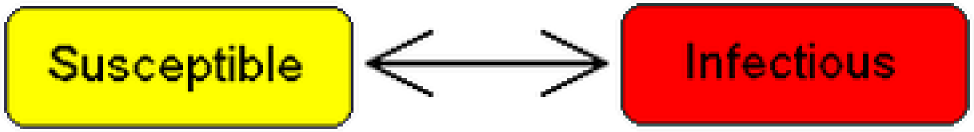
R(t) : #recovered people at time t;

β *: a parameter for infectivity;*

γ *: a parameter for recovery.*

SIS Model

- Designed for infections confer no long lasting immunity (e.g., common cold)
- Individuals are considered become susceptible again after infection:



- Model:

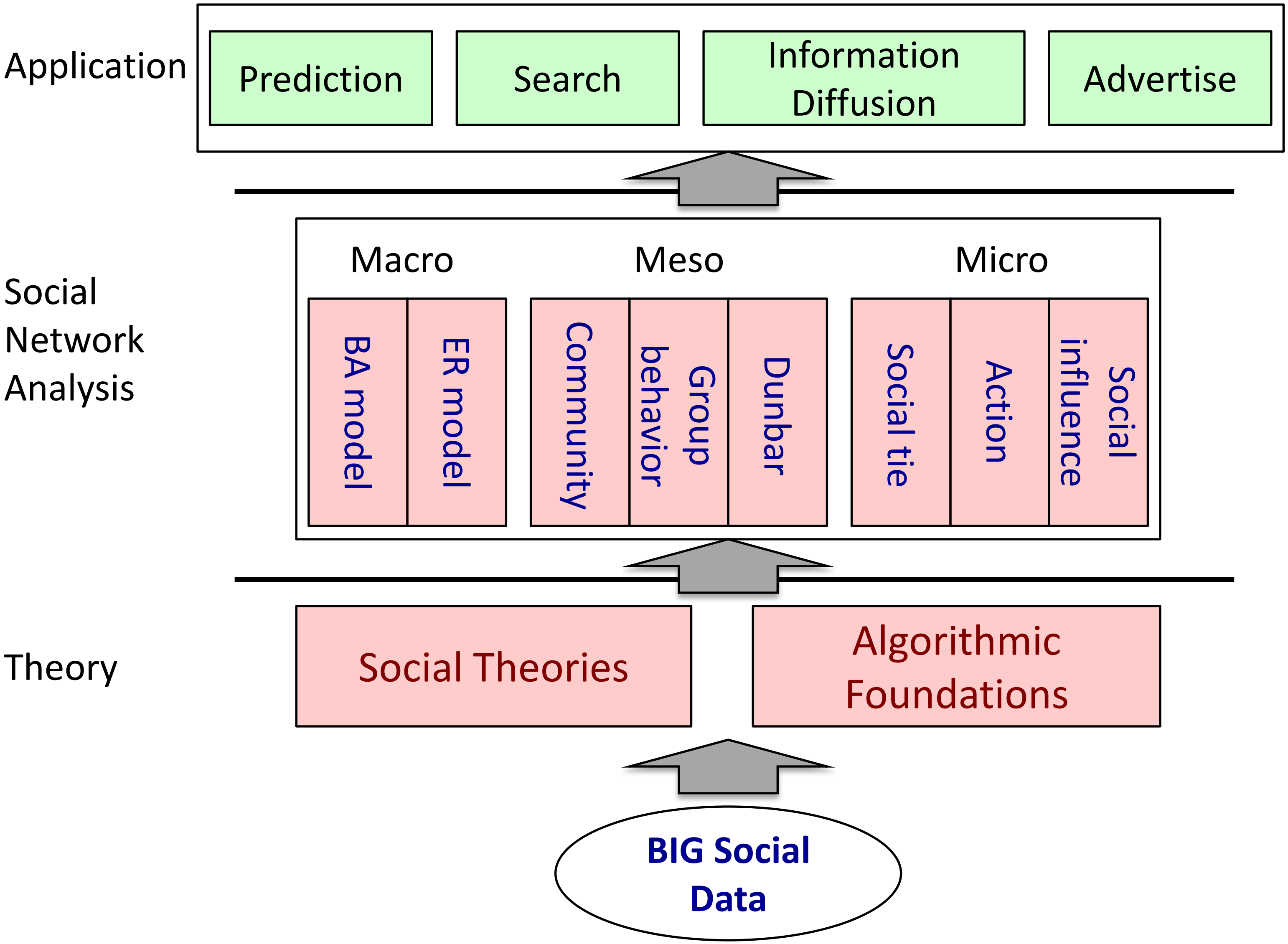
$$\frac{dS}{dt} = -\beta SI + \gamma I$$
$$\frac{dI}{dt} = \beta SI - \gamma I$$

Notice for both SIR and SIS, it holds:

$$\frac{dS}{dt} + \frac{dI}{dt} = 0 \Rightarrow S(t) + I(t) = N$$

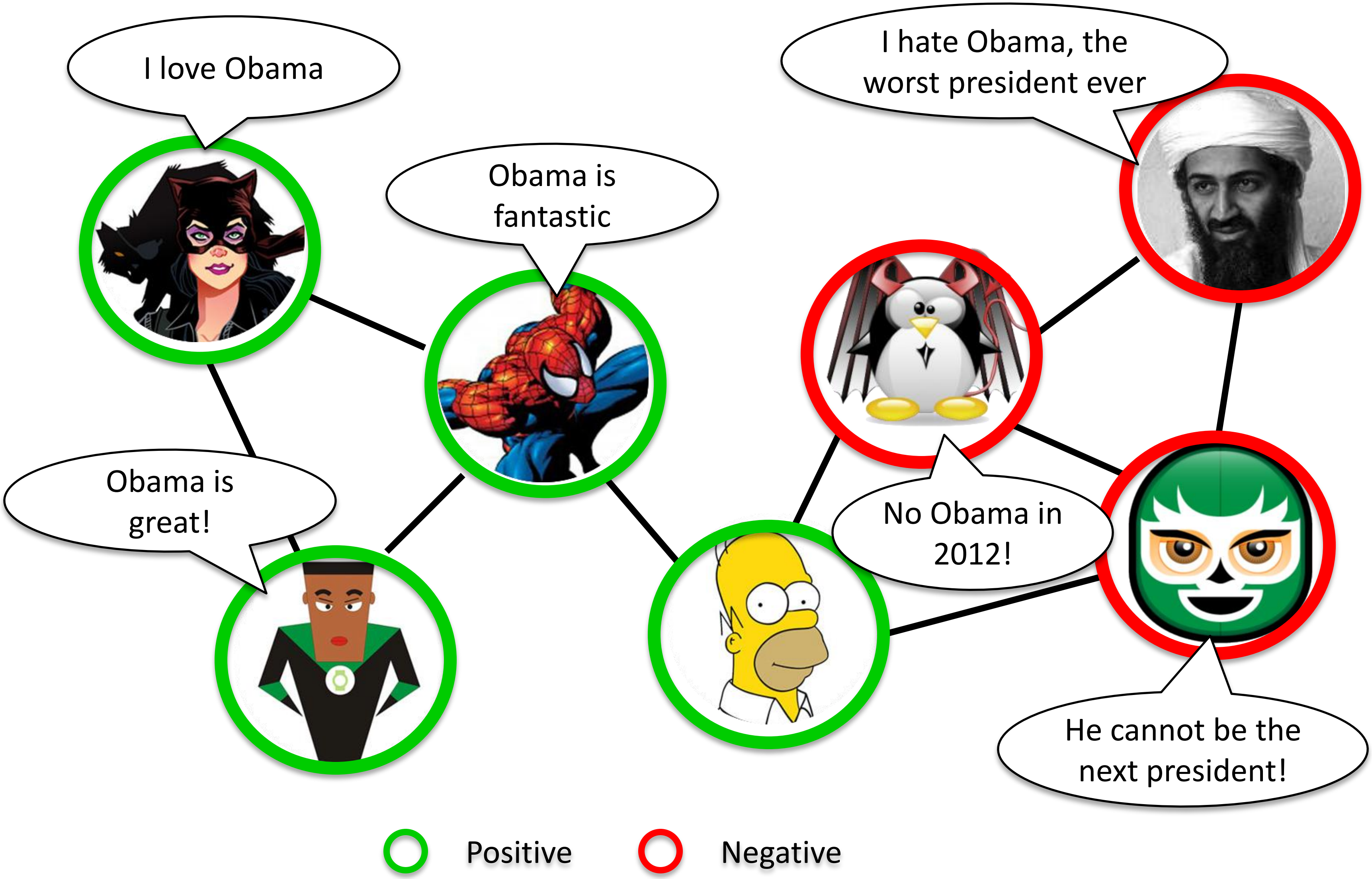
where N is the ***fixed*** *total population.*

Core Research in Social Network



Social Influence Analysis

“Love Obama”


I love Obama
Obama is fantastic
I hate Obama, the worst president ever
Obama is great!
No Obama in 2012!
He cannot be the next president!
○ Positive
○ Negative

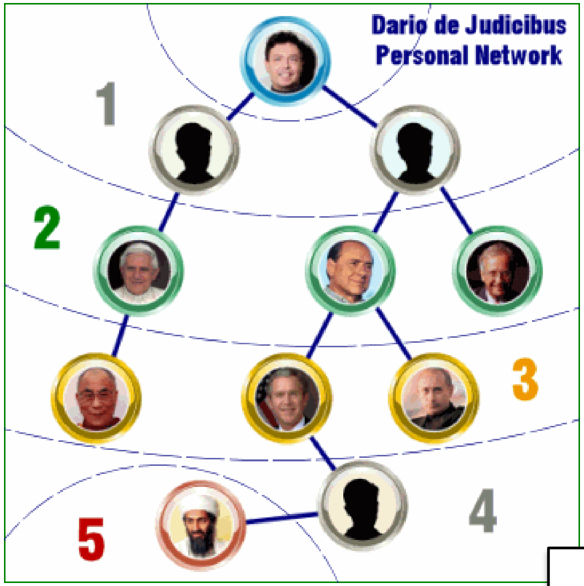
What is Social Influence?

- Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally.[1]
  - **Informational social influence**: to accept information from another;
  - **Normative social influence**: to conform to the positive expectations of others.

[1] http://en.wikipedia.org/wiki/Social_influence

Three Degree of Influence

Six degree of separation[1]



Three degree of Influence[2]

You are able to **influence** up to >1,000,000 persons in the world, according to the Dunbar's number[3].

[1] S. Milgram. The Small World Problem. Psychology Today, 1967, Vol. 2, 60–67

[2] J.H. Fowler and N.A. Christakis. The Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study. British Medical Journal 2008; 337: a2338

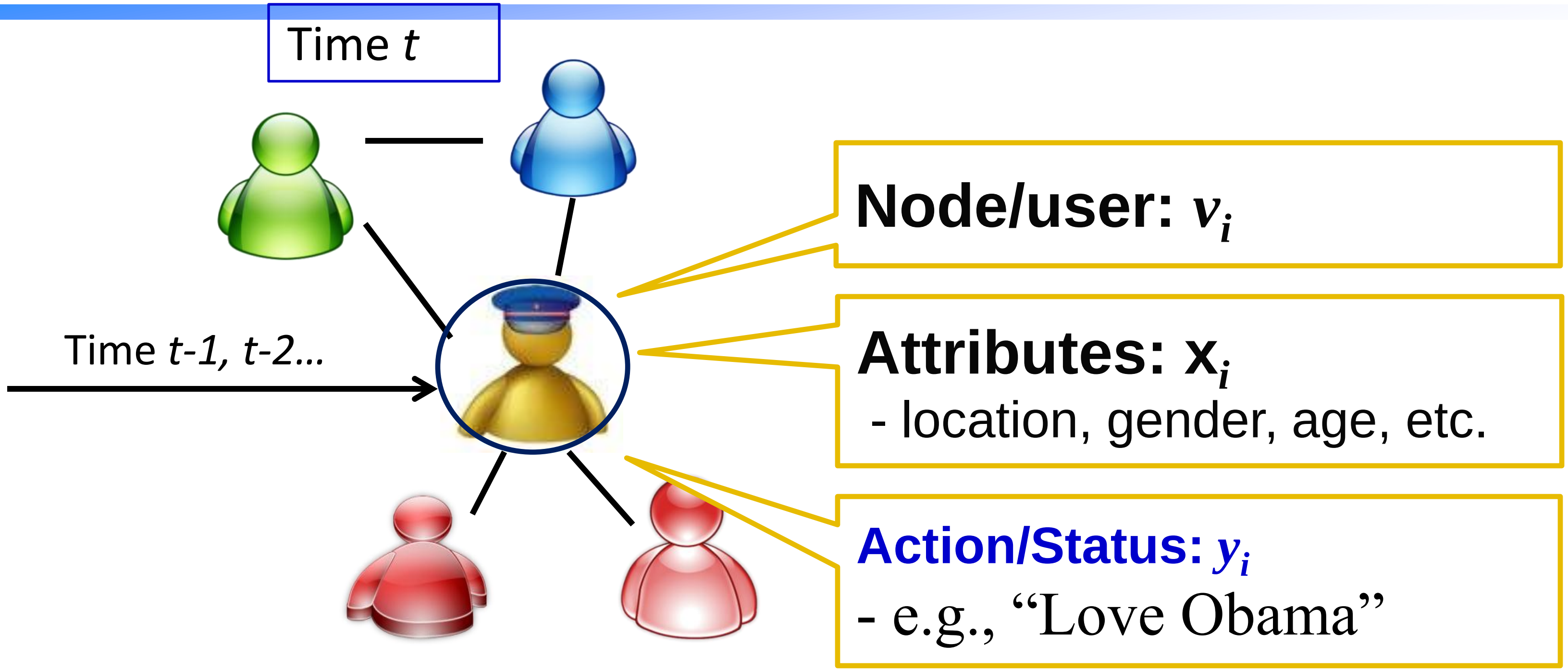
[3] R. Dunbar. Neocortex size as a constraint on group size in primates. Human Evolution, 1992, 20: 469–493.

Challenges: WH^3

1. Whether social influence exist?
2. How to measure influence?
3. How to model influence?
4. How influence can help real applications?

Preliminaries

Notations



Node/user: v_i

Attributes: $\mathbf{x}_i$

- location, gender, age, etc.

Action/Status: y_i

- e.g., “Love Obama”

$$G = (V, E, X, Y)$$

G^t — the superscript *t* represents the time stamp

e_{ij}^{t} Î E^{t} — represents a link/relationship from v_i to v_j at time *t*

Homophily

- Homophily
  - A user in the social network tends to be similar to their connected neighbors.
- Originated from different mechanisms
  - Social influence
    - Indicates people tend to follow the behaviors of their friends
  - Selection
    - Indicates people tend to create relationships with other people who are already similar to them
  - Confounding variables
    - Other unknown variables exist, which may cause friends to behave similarly with one another.

Influence and Selection[1]

$$Selection = \frac{p(e_{ij}^{t} = 1 \mid e_{ij}^{t-1} = 0, \boxed{\left\langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \right\rangle > e})}{p(\boxed{e_{ij}^{t} = 1} \mid e_{ij}^{t-1} = 0)}$$

Similarity between user *i* and *j* at time *t*-1 is larger than a threshold

There is a link between user *i* and *j* at time *t*

- Denominator: the conditional probability that an unlinked pair will become linked
- Numerator: the same probability for unlinked pairs whose similarity exceeds the threshold

$$Influence = \frac{p(\left\langle \mathbf{x}_i^{t}, \mathbf{x}_j^{t} \right\rangle > \left\langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \right\rangle \mid e_{ij}^{t} = 1, e_{ij}^{t-1} = 0)}{p(\left\langle \mathbf{x}_i^{t}, \mathbf{x}_j^{t} \right\rangle > \left\langle \mathbf{x}_i^{t-1}, \mathbf{x}_j^{t-1} \right\rangle \mid e_{ij}^{t-1} = 0)}$$

- Denominator: the probability that the similarity increase from time *t*-1 to time *t* between two nodes that were not linked at time *t*-1
- Numerator: the same probability that became linked at time *t*

- A Model is learned through matrix factorization/factor graph

[1] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. In KDD'09, pages 747–756, 2009.

Other Related Concepts

- Cosine similarity
- Correlation factors
- Hazard ratio
- *t*-test

Cosine Similarity

- A measure of similarity
- Use a vector to represent a sample (e.g., user)

$$\mathbf{x} = (x_1, \ldots, x_n)$$

- To measure the similarity of two vectors **x** and **y**, employ cosine similarity:

$$sim(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \times \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

Correlation Factors

- Several correlation coefficients could be used to measure correlation between two random variables *x* and *y*.
- Pearsons' correlation

$$r_{x,y} = corr(x, y) = \frac{E[(x - m_x)(y - \boxed{m_y})]}{s_x \boxed{s_y}}$$

mean

Standard deviation

- It could be estimated by

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

- Note that **correlation** does NOT imply **causation**

Hazard Ratio

- **Hazard Ratio**
  - Chance of an event occurring in the treatment group divided by its chance in the control group
  - Example:

$$\frac{\text{Chance of users to buy iPhone with >=1 iPhone user friend(s)}}{\text{Chance of users to buy iPhone without any iPhone user friend}}$$

  - Measuring instantaneous chance by *hazard rate h(t)*

$$h(t) = \lim_{\Delta t \to 0} \frac{\text{observed events in interval}[t, t + \Delta t]/N(t)}{\Delta t}$$

  - The hazard ratio is the relationship between the instantaneous hazards in two groups
  - Proportional hazards models (e.g. Cox-model) could be used to report hazard ratio.

t-test

- A *t*-test usually used when the test statistic follows a Student's *t* distribution if the null hypothesis is supported.
- To test if the difference between two variables are significant
- Welch's *t*-test
  - Calculate *t*-value

sample mean

Unbiased estimator of sample variance

$$t = \frac{\boxed{\overline{x}_1} - \overline{x}_2}{s_{\overline{x}_1 - \overline{x}_2}}, \quad s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\boxed{s_1^2}}{\boxed{n_1}} + \frac{s_2^2}{\boxed{n_2}}}$$

#participants in the control group

#participants in the treatment group

  - Find the *p*-value using a table of values from Student's t-distribution
  - If the *p*-value is below chosen threshold (e.g. 0.01) then the two variables are viewed as significant different.

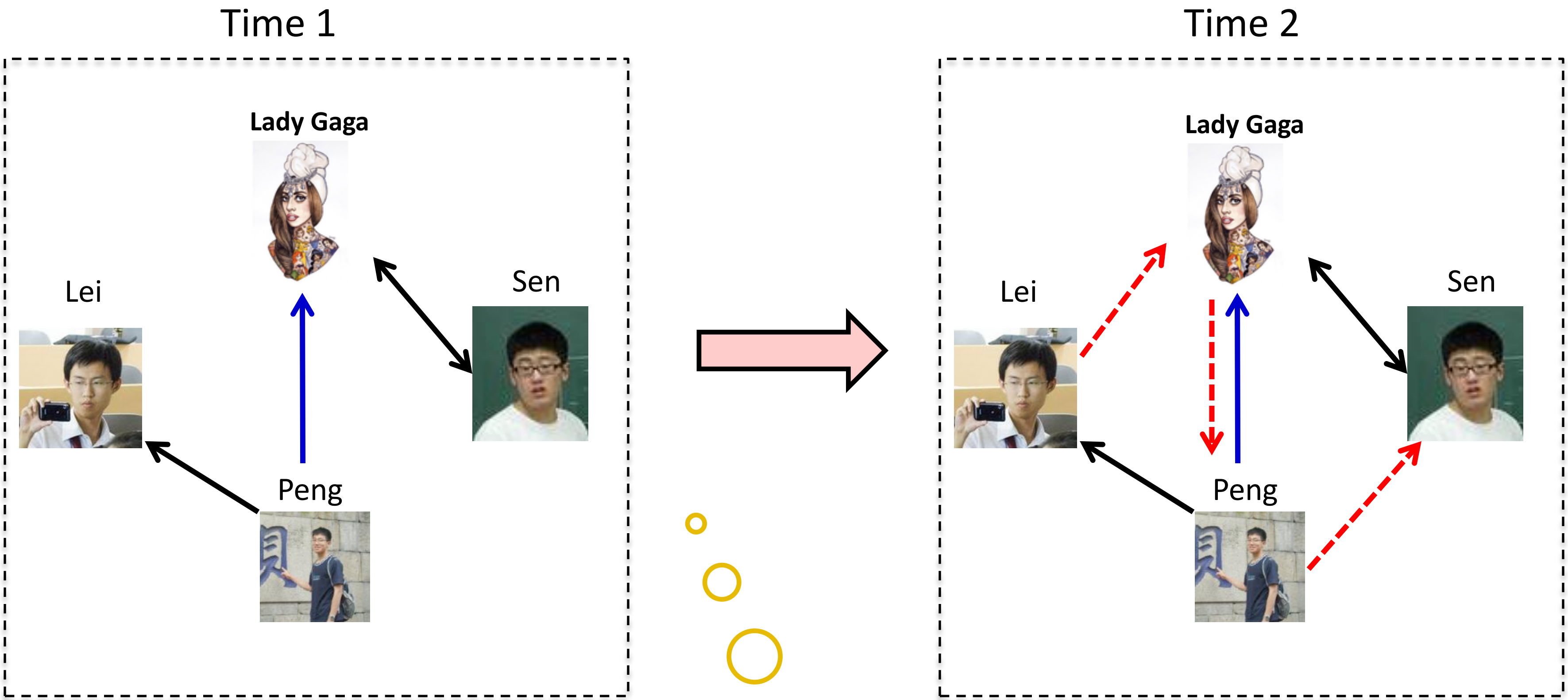
Data Sets

Ten Cases

Network	#Nodes	#Edges	Behavior
Twitter-net	111,000	450,000	Follow
Weibo-Retweet	1,700,000	400,000,000	Retweet
Slashdot	93,133	964,562	Friend/Foe
Mobile (THU)	229	29,136	Happy/Unhappy
Gowalla	196,591	950,327	Check-in
ArnetMiner	1,300,000	23,003,231	Publish on a topic
Flickr	1,991,509	208,118,719	Join a group
PatentMiner	4,000,000	32,000,000	Patent on a topic
Citation	1,572,277	2,084,019	Cite a paper
Twitter-content	7,521	304,275	Tweet “Haiti Earthquake”

Most of the data sets will be publicly available for research.

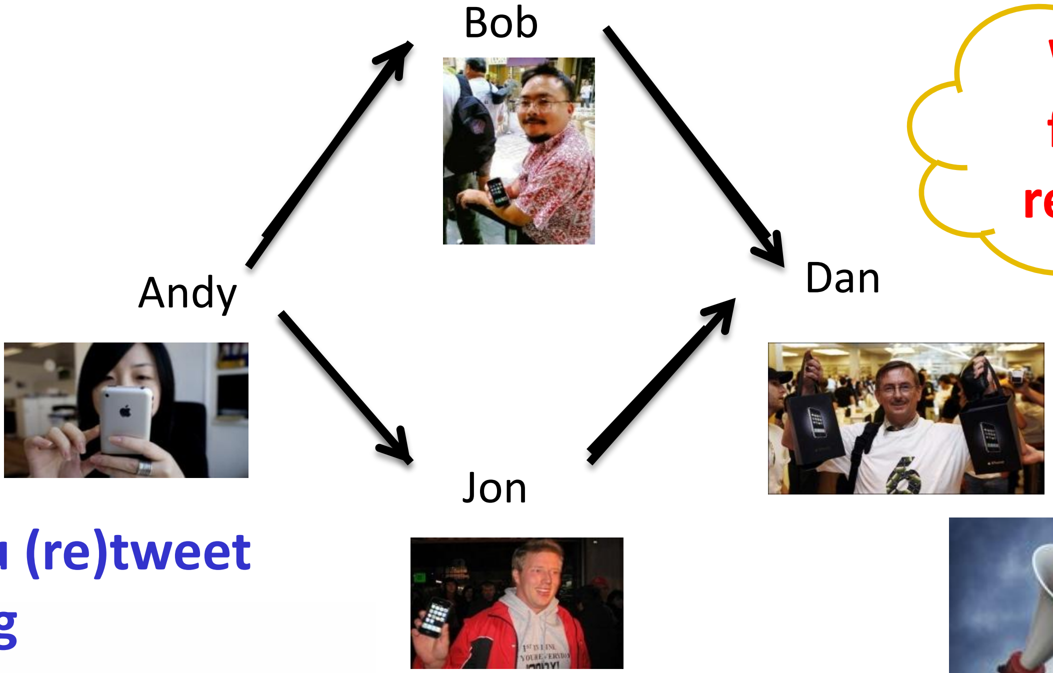
Case 1: Following Influence on Twitter



When you follow a user in a social network, will the behavior influences your friends to also follow her?

Case 2: Retweeting Influence

新浪微博
weibo.com



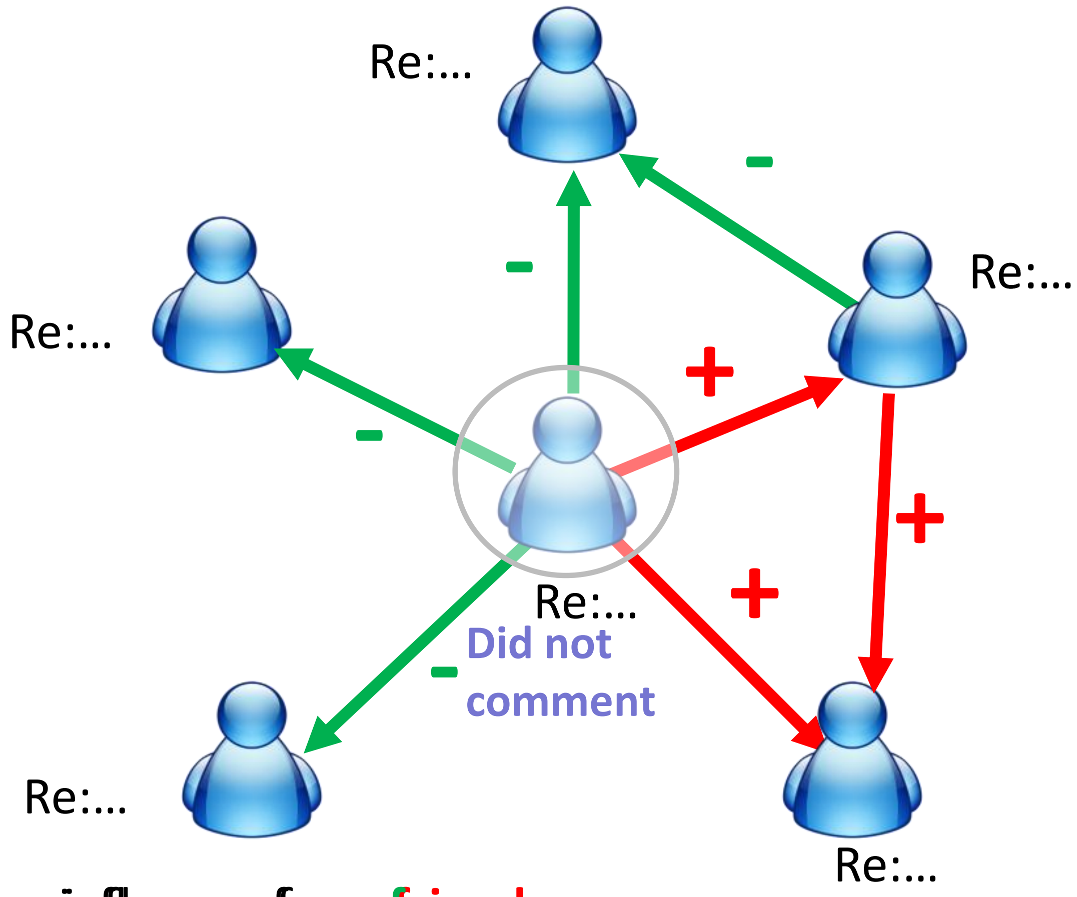
Who will follow to retweet it?

When you (re)tweet something

Case 3: Commenting Influence

News: Alan Cox Exits Intel.

Governments Want Private Data



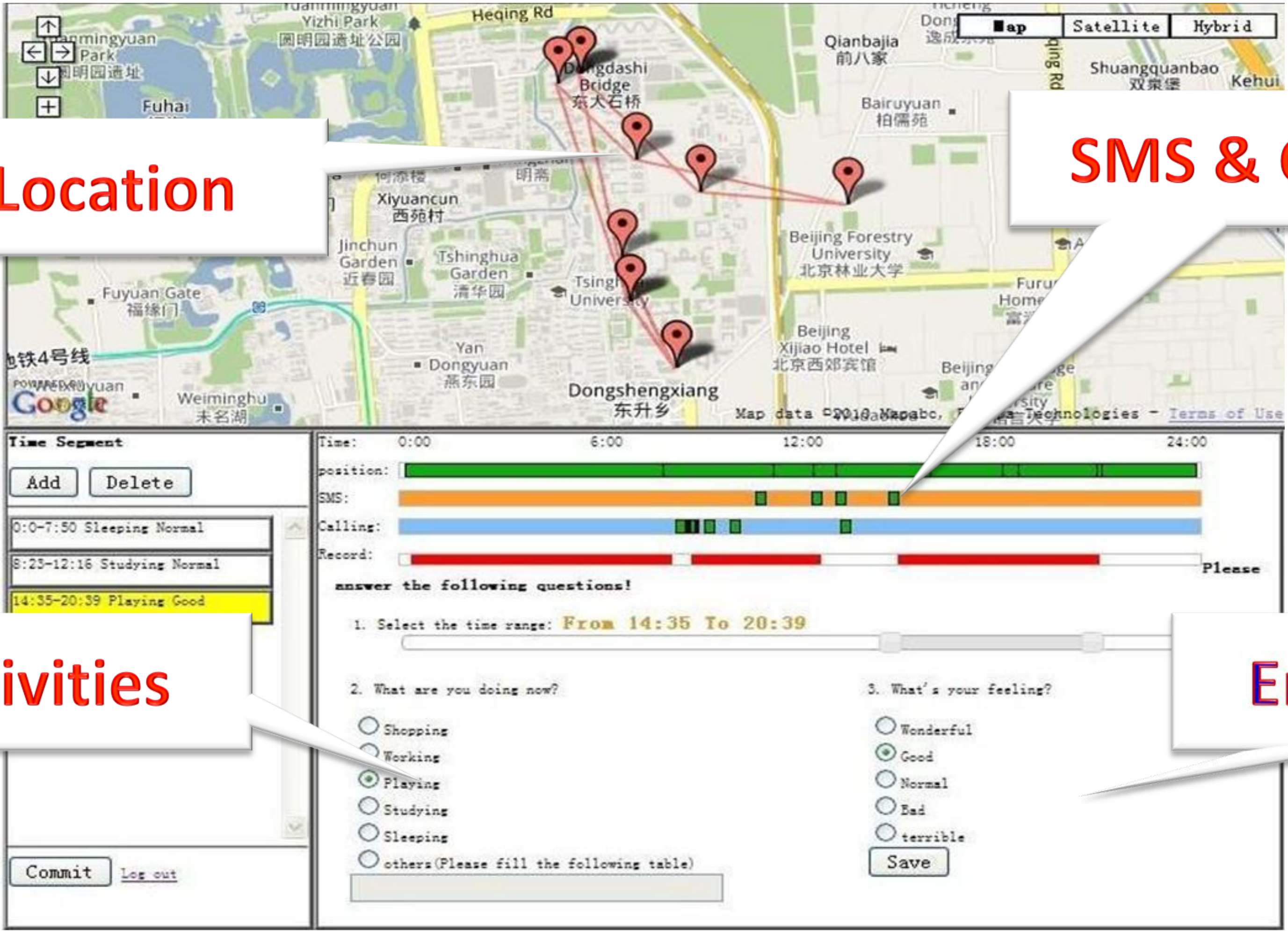
+ Friend

- Foe

negative influence from foes

positive influence from friends

Case 4: Emotion Influence



Time Segment

Add Delete

0:0-7:50 Sleeping Normal

8:23-12:16 Studying Normal

14:35-20:39 Playing Good

Commit Log out

Please answer the following questions!

1. Select the time range: From 14:35 To 20:39

2. What are you doing now?

- ○ Shopping
- ○ Working
- ● Playing
- ○ Studying
- ○ Sleeping
- ○ others(Please fill the following table)

3. What's your feeling?

- ○ Wonderful
- ● Good
- ○ Normal
- ○ Bad
- ○ terrible

Save

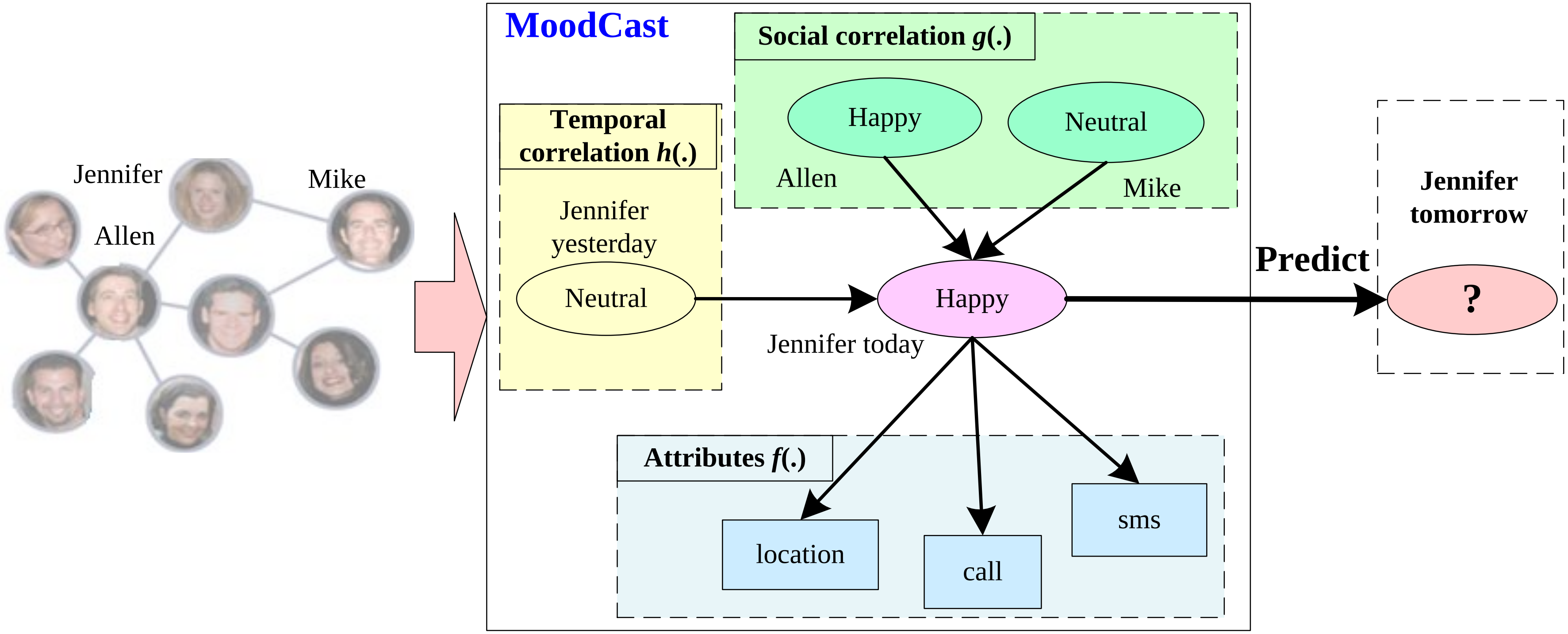
Location

SMS & Calling

Activities

Emotion?

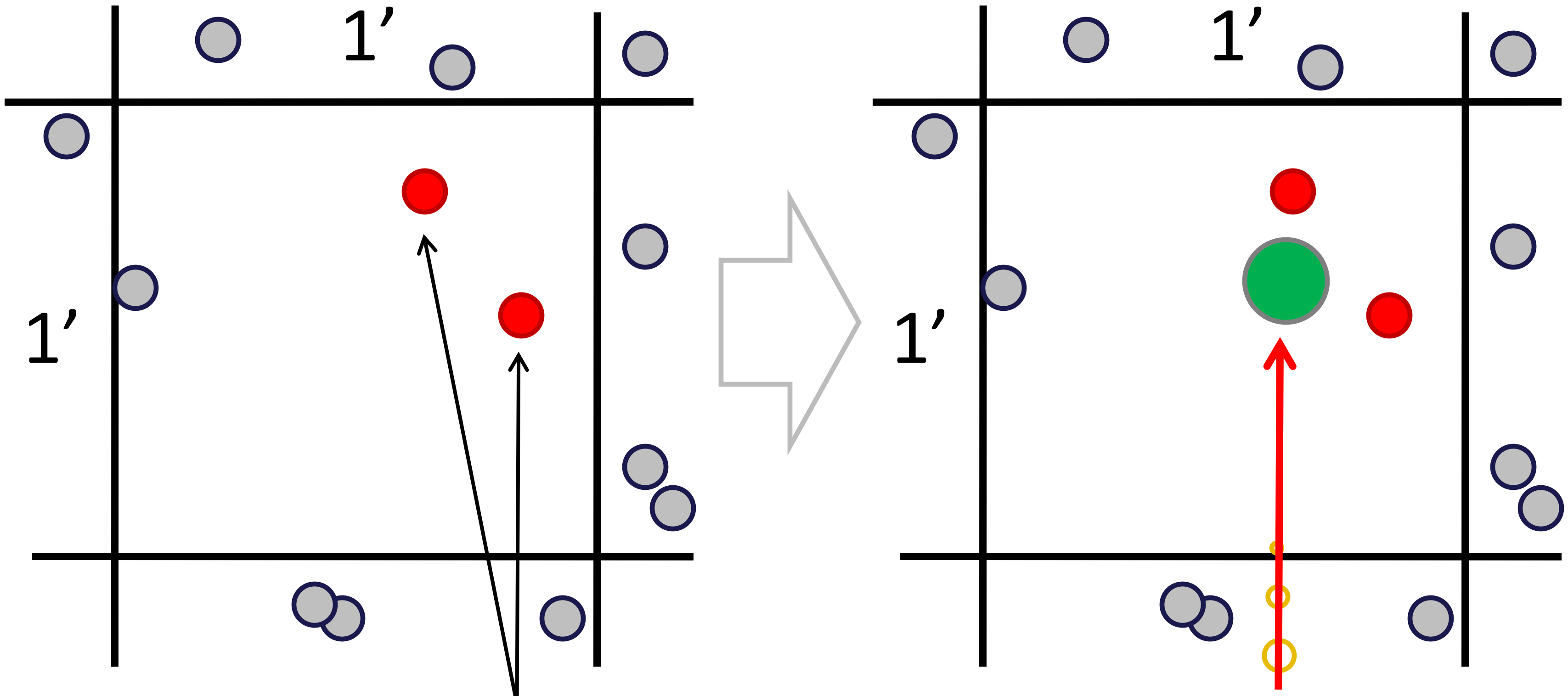
Case 4: Emotion Influence (cont.)



Can we predict users' emotion?

Case 5: Check-in Influence in Gowalla

Legend 🟢 Alice 🔴 Alice’s friend ⚪ Other users



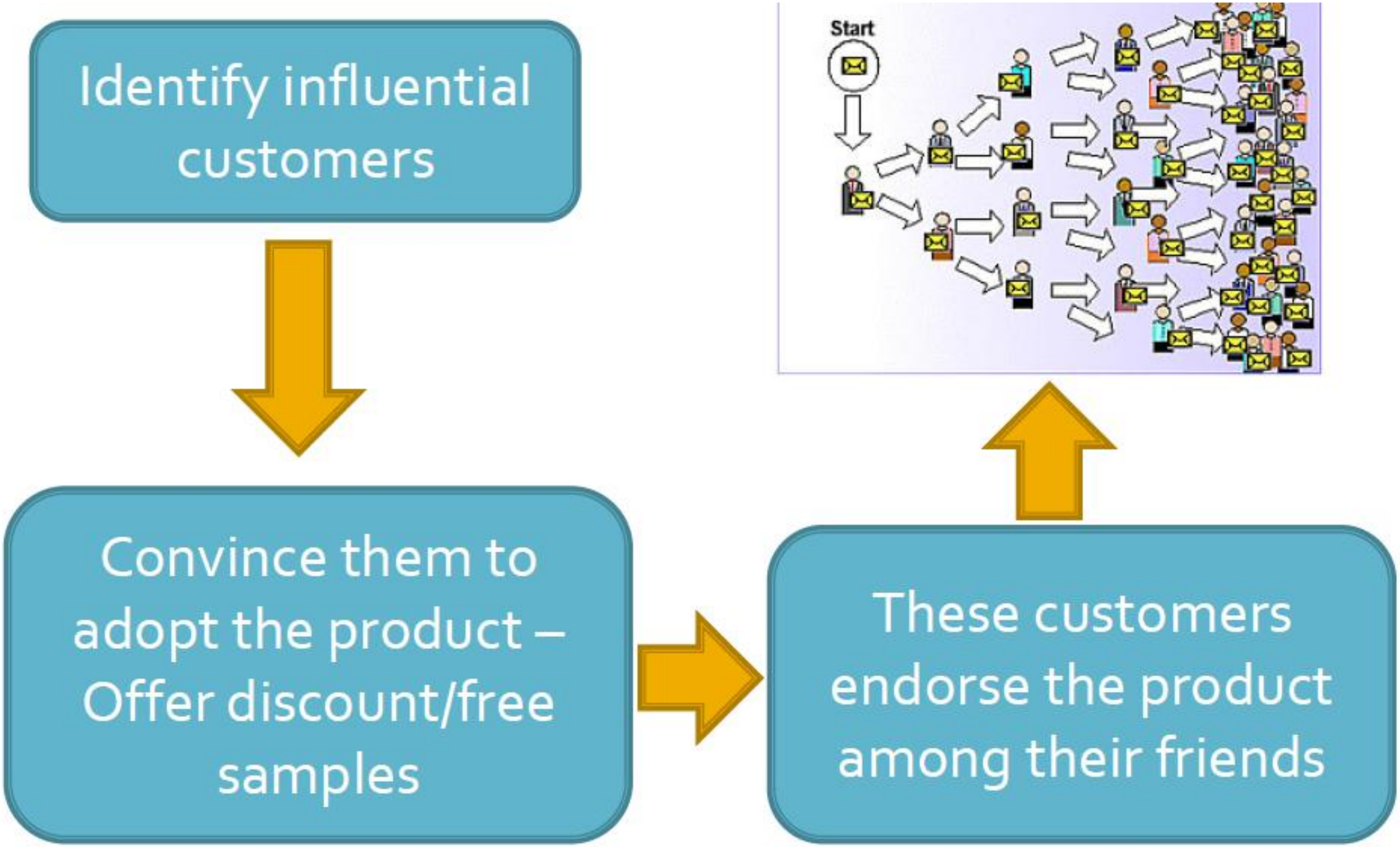
If Alice’s friends check in this location at time *t*

Will Alice also check in nearby?

Understanding the Emotional Impact in Social Networks

[1] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can We Understand van Gogh's Mood? Learning to Infer Affects from Images in Social Networks. In ACM Multimedia, pages 857-860, 2012.

The model of Viral Marketing



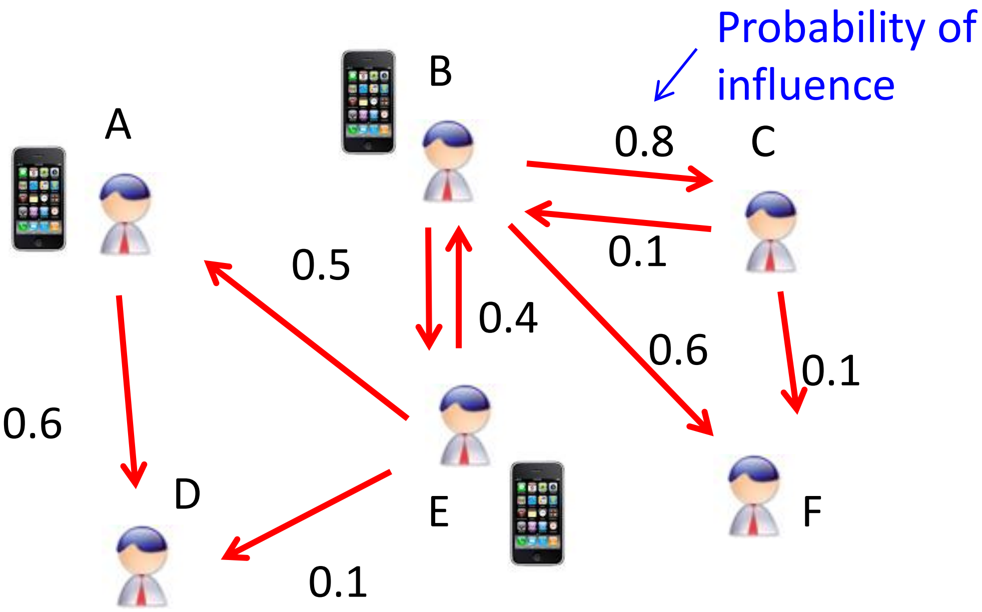
Identify influential customers

Convince them to adopt the product – Offer discount/free samples

These customers endorse the product among their friends

Influence Maximization

- Influence maximization
  - Minimize marketing cost and more generally to maximize profit.
  - E.g., to get a small number of influential users to adopt a new product, and subsequently trigger a large cascade of further adoptions.



[1] P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'01), pages 57–66, 2001.

Problem Abstraction

- We associate each user with a status:
  - Active or Inactive
  - The status of the chosen set of users (seed nodes) to market is viewed as active
  - Other users are viewed as inactive
- Influence maximization
  - Initially all users are considered inactive
  - Then the chosen users are activated, who may further influence their friends to be active as well

Diffusion Influence Model

- Linear Threshold Model
- Cascade Model

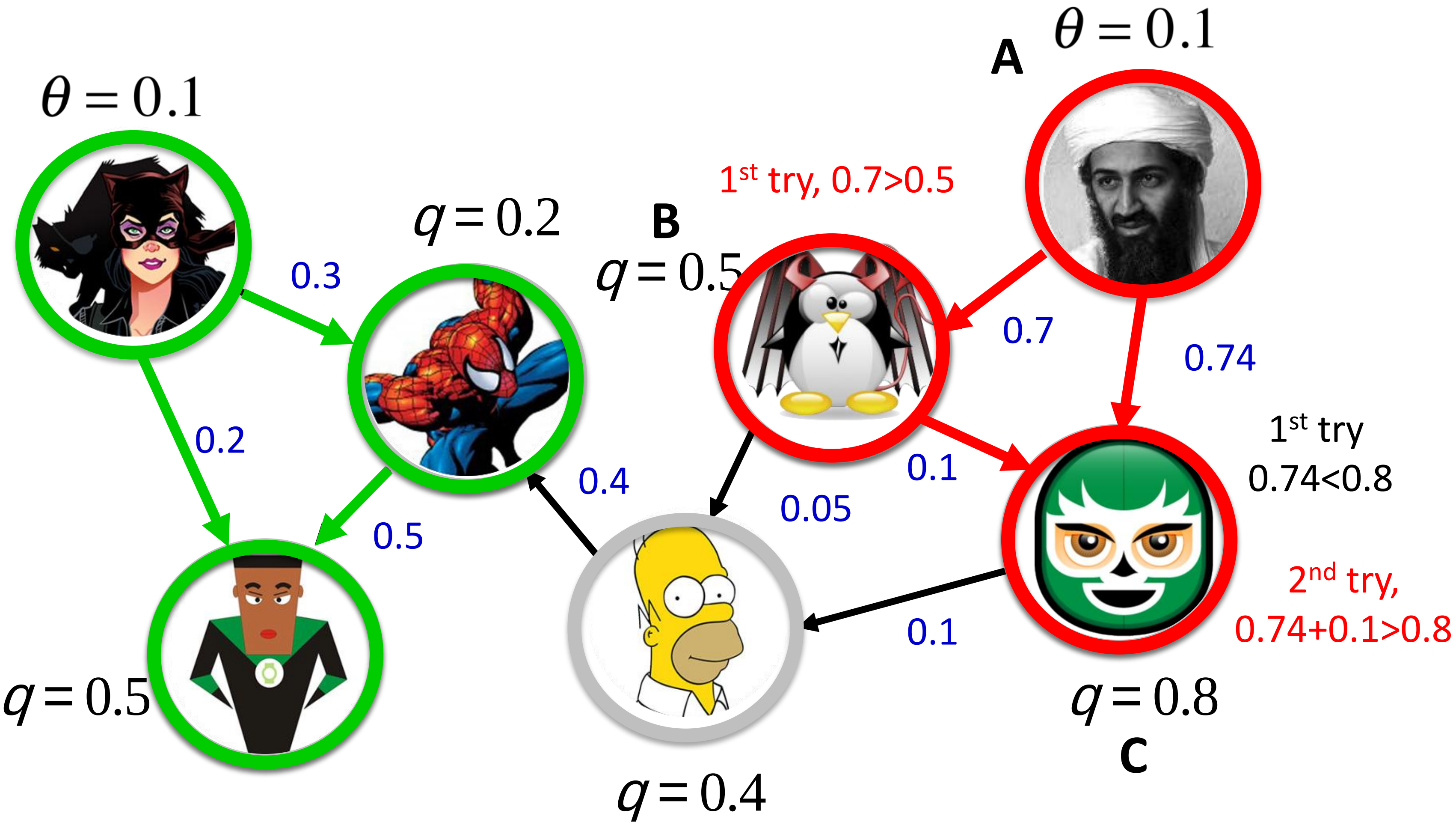
Linear Threshold Model

- General idea
  - Whether a given node will be active can be based on an arbitrary monotone function of its neighbors that are already active.
- Formalization
  - f_v : map subsets of *v*'s neighbors' influence to real numbers in [0,1]
  - θ_v : a threshold for each node
  - *S*: the set of neighbors of *v* that are active in step *t*-1
  - Node *v* will turn active in step *t* if $f_v(S) > \theta_v$
- Specifically, in [Kempe, 2003], f_v is defined $\sum_{u \in S} b_{v.u}$, where $b_{v,u}$ can be seen as a fixed weight, satisfying

$$\sum_{v \in N(u)} b_{u,v} \leq 1$$

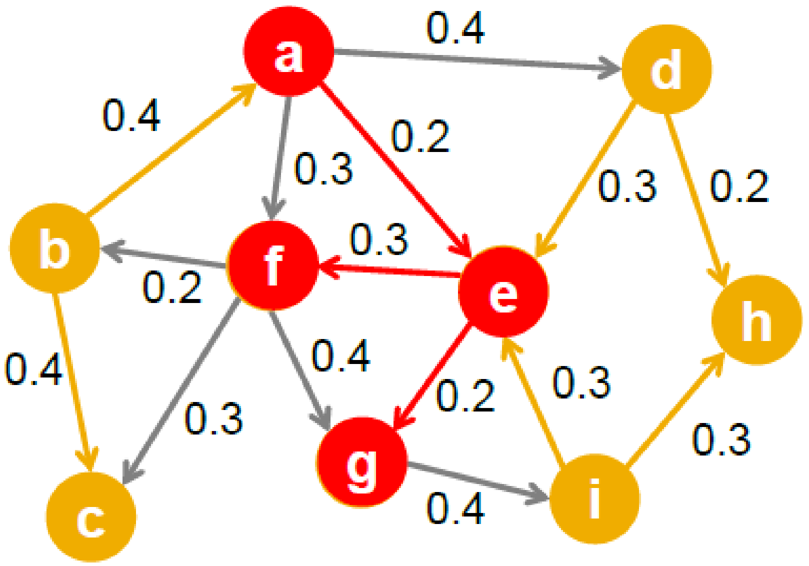
[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'03), pages 137–146, 2003.

Linear Threshold Model: An example



Independent Cascade model

- Initially some nodes *S* are active
- Each edge (v, w) has probability (weight) p_{vw}



- When node v becomes active:
  - It activates each out-neighbor w with prob. p_{vw}
- Activations spread through the network

Cascade Model

- Cascade model
  - $p_v(u,S)$: the success probability of user *u* activating user v
  - User *u* tries to activate v and finally succeeds, where *S* is the set of *v*'s neighbors that have already attempted but failed to make *v* active
- Independent cascade model
  - $p_v(u,S)$ is a constant, meaning that whether *v* is to be active does not depend on the order *v*'s neighbors try to activate it.
  - Key idea: Flip coins *c* in advance -> live edges
  - $F_c(A)$: People influenced under outcome *c* (set cover)
  - $F(A)$ = Sum $_c$P(c) $F_c(A)$ is submodular as well

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'03), pages 137–146, 2003.

Theoretical Analysis

- NP-hard [1]
  - Linear threshold model
  - General cascade model
- Kempe Prove that approximation algorithms can guarantee that the influence spread is within(1-1/e) of the optimal influence spread.
  - Verify that the two models can outperform the traditional heuristics
- Recent research focuses on the efficiency improvement
  - [2] accelerate the influence procedure by up to 700 times
- It is still challenging to extend these methods to large data sets

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining(KDD'03), pages 137–146, 2003.
[2] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'07), pages 420–429, 2007.

Objective Function

- **Objective function:**

  - *f* (*S*) = Expected #people influenced when targeting a set of users *S*

- Define *f* (*S*) as a monotonic submodular function

$$f(S \cup \{v\}) - f(S) \quad \geq \quad f(T \cup \{v\}) - f(T)$$

$$f(S \cup \{v\}) \geq f(S)$$

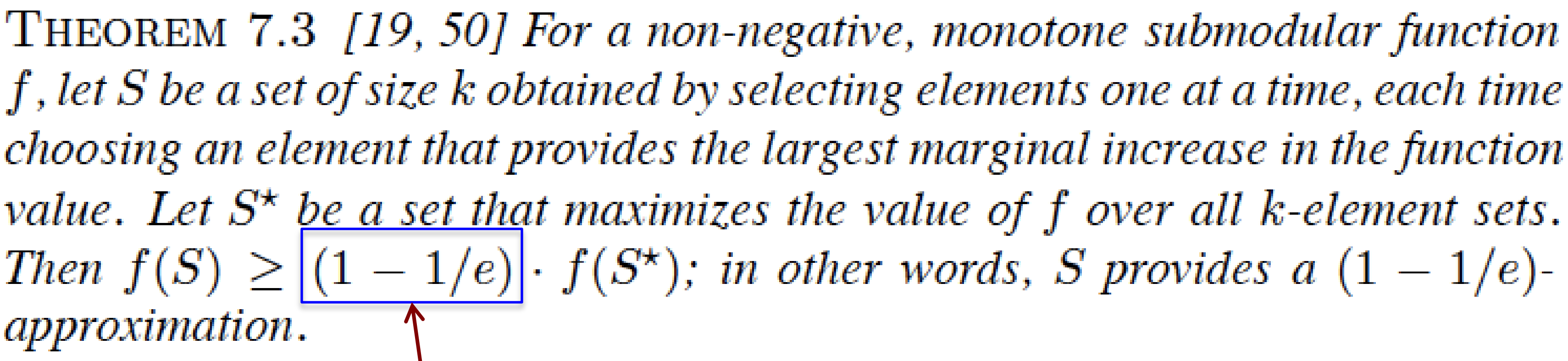
where $S \subseteq T$.

[1] P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'01), pages 57–66, 2001.

[2] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining(KDD'03), pages 137–146, 2003.

Maximizing the Spread of Influence

- Solution
  - Use a submodular function to approximate the influence function
  - Then the problem can be transformed into finding a *k*-element set *S* for which *f* (*S*) is maximized.



THEOREM 7.3 *[19, 50] For a non-negative, monotone submodular function f, let S be a set of size k obtained by selecting elements one at a time, each time choosing an element that provides the largest marginal increase in the function value. Let $S^\star$ be a set that maximizes the value of f over all k-element sets. Then $f(S) \geq \boxed{(1-1/e)} \cdot f(S^\star)$; in other words, S provides a $(1-1/e)$-approximation.*

approximation ratio

[1] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'03), pages 137–146, 2003.

Performance Guarantee

Let g_j be the j-th node selected by the greedy algorithm

- Let $G_j = \{g_1, \ldots, g_j\}$ and $G_0 = \emptyset$
- For $\forall S, |S| = k$ and $j = 0, 1, \ldots, k-1$

$$F(S) \underset{\substack{\uparrow \\ \textbf{monotonicity}}}{\le} F(G_j \cup S) \underset{\substack{\uparrow \\ \textbf{greedy +} \\ \textbf{submodularity}}}{\le} F(G_j) + k g_{j+1}$$

- Let $\Delta_j = F(S^*) - F(G_j)$

where S^* is the optimal solution

- We have $g_{j+1} = \Delta_j - \Delta_{j+1}$

- Thus $\Delta_j \le k(\Delta_j - \Delta_{j+1})$

$$\Delta_k \le \left(1 - \frac{1}{k}\right)^k \Delta_0$$

$$\le \frac{1}{e} F(S^*)$$

Recall
$e^x \ge 1 + x$

- Then

$$F(G_k) \ge \left(1 - \frac{1}{e}\right) F(S^*)$$

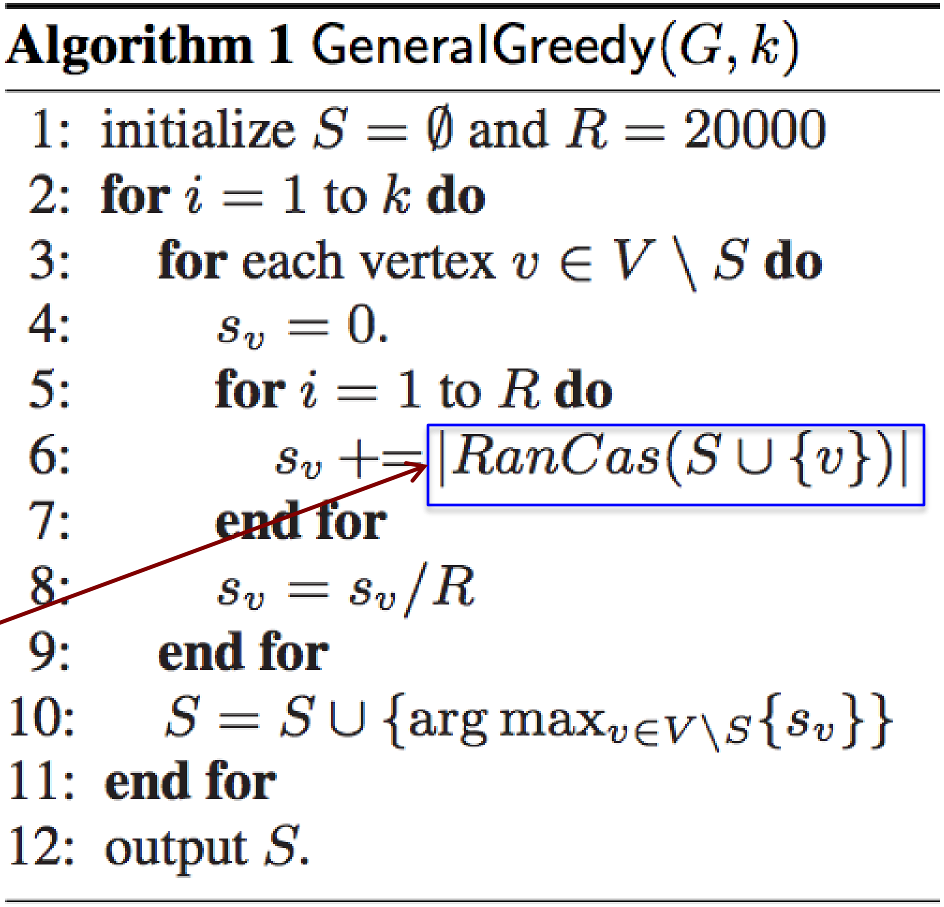
*The solution obtained by Greedy is better than **63%** of the optimal solution*

Algorithms

- General Greedy
- Low-distance Heuristic
- High-degree heuristic
- Degree Discount Heuristic

General Greedy

- General idea: In each round, the algorithm adds one vertex into the selected set *S* such that this vertex together with current set *S* maximizes the influence spread.



Algorithm 1 GeneralGreedy(G, k)

```
initialize $S = \emptyset$ and $R = 20000$
for $i = 1$ to $k$ do
    for each vertex $v \in V \setminus S$ do
        $s_v = 0$.
        for $i = 1$ to $R$ do
            $s_v \mathrel{+}= \boxed{|RanCas(S \cup \{v\})|}$
        end for
        $s_v = s_v / R$
    end for
    $S = S \cup \{\arg\max_{v \in V \setminus S}\{s_v\}\}$
end for
output $S$.
```

Any random diffusion process

Low-distance Heuristic

- Consider the nodes with the shortest paths to other nodes as seed nodes
- Intuition
  - Individuals are more likely to be influenced by those who are closely related to them.

High-degree heuristic

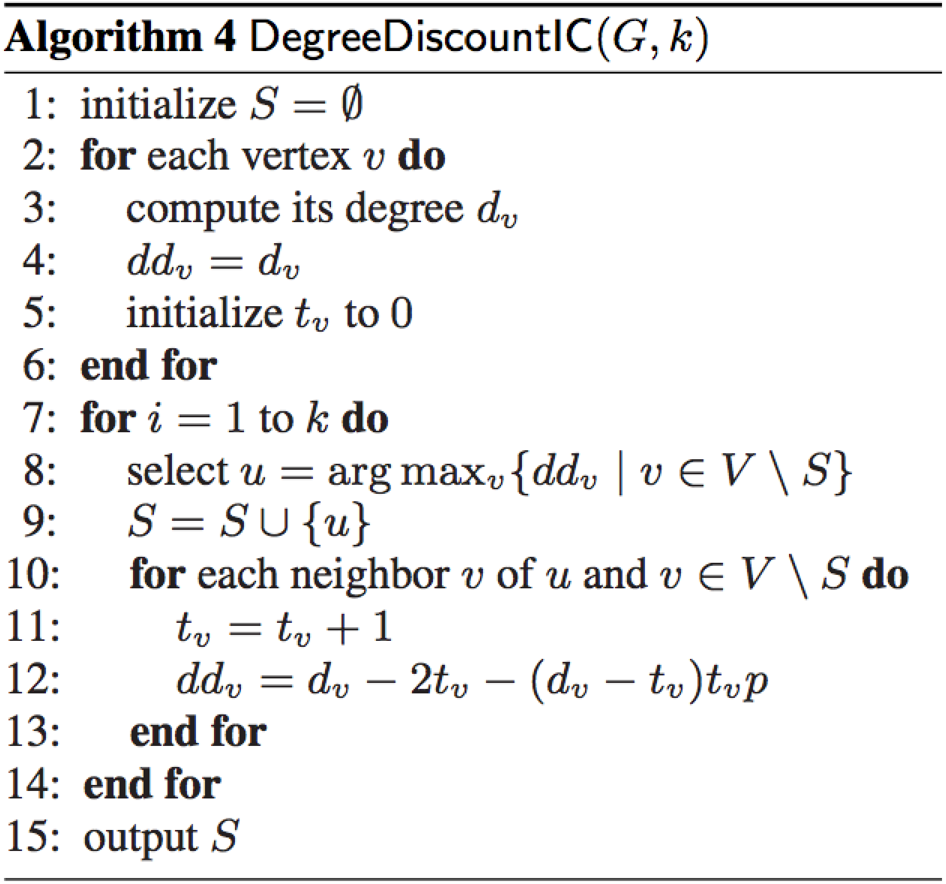
- Choose the seed nodes according to their degree.
- Intuition
  - The nodes with more neighbors would arguably tend to impose more influence upon its direct neighbors.
  - Know as “degree centrality”

Degree Discount Heuristic[1]

- General idea: If *u* has been selected as a seed, then when considering selecting *v* as a new seed based on its degree, we should not count the edge *v->u*
- Specifically, for a node *v* with d_v neighbors of which t_v are selected as seeds, we should discount *v*'s degree by

$$2t_v + (d_v - t_v)\, t_v p$$

where p=0.1.

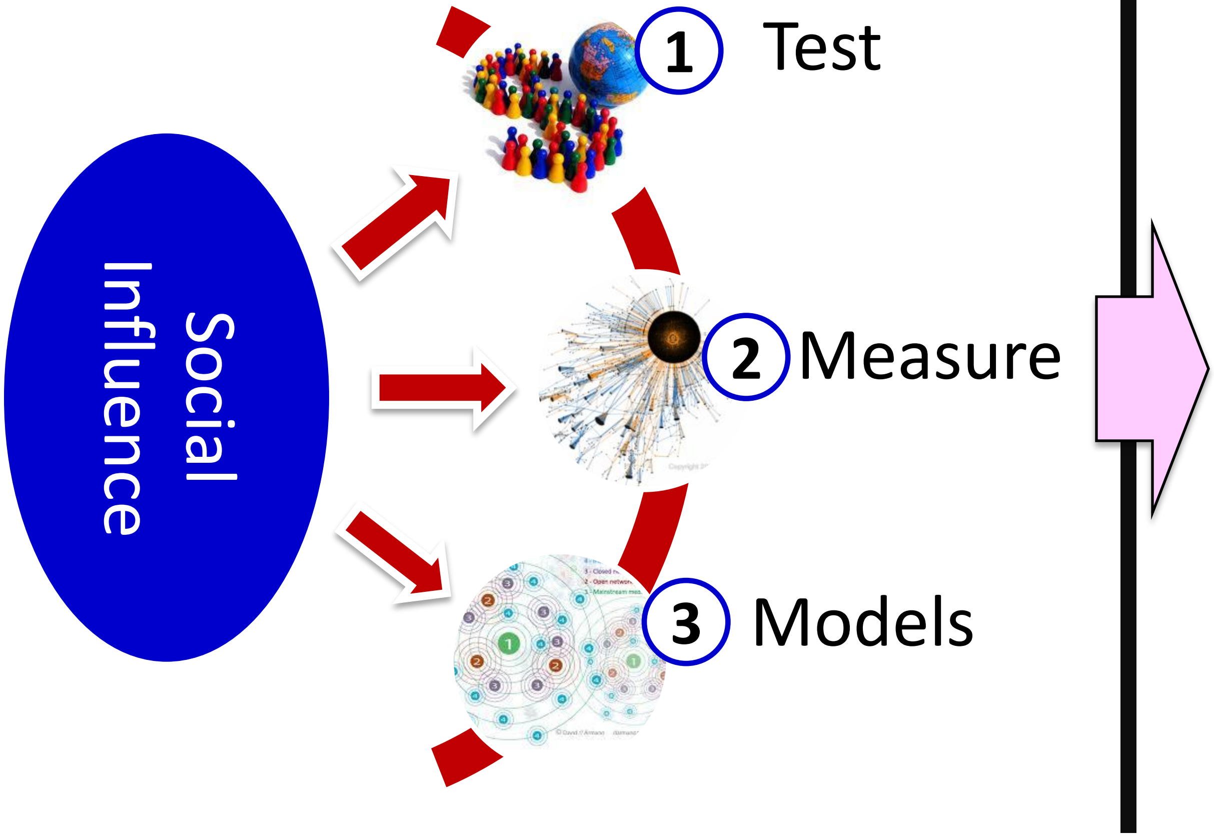


Algorithm 4 DegreeDiscountIC(G, k)

```
initialize $S = \emptyset$
for each vertex $v$ do
    compute its degree $d_v$
    $dd_v = d_v$
    initialize $t_v$ to 0
end for
for $i = 1$ to $k$ do
    select $u = \arg\max_v \{dd_v \mid v \in V \setminus S\}$
    $S = S \cup \{u\}$
    for each neighbor $v$ of $u$ and $v \in V \setminus S$ do
        $t_v = t_v + 1$
        $dd_v = d_v - 2t_v - (d_v - t_v)t_v p$
    end for
end for
output $S$
```

[1] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In KDD'09, pages 199-207, 2009.

Social Influence



Applications

Application: Social Advertising[1]

- Conducted two very large field experiments that identify the effect of social cues on consumer responses to ads on Facebook
- Exp. 1: measure how responses increase as a function of the number of cues.
- Exp. 2: examines the effect of augmenting traditional ad units with a minimal social cue
- Result: Social influence causes significant increases in ad performance

[1] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn. Social influence in social advertising: evidence from field experiments. In EC'12, pages 146-161, 2012.

Application: Opinion Leader[1]

- Propose viral marketing through frequent pattern mining.
- Assumption
  - Users can see their friends actions.
- Basic formation of the problem
  - Actions take place in different time steps, and the actions which come up later could be influenced by the earlier taken actions.
- Approach
  - Define leaders as people who can influence a sufficient number of people in the network with their actions for a long enough period of time.
  - Finding leaders in a social network makes use of action logs.

[1] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In CIKM’08, pages 499–508, 2008.

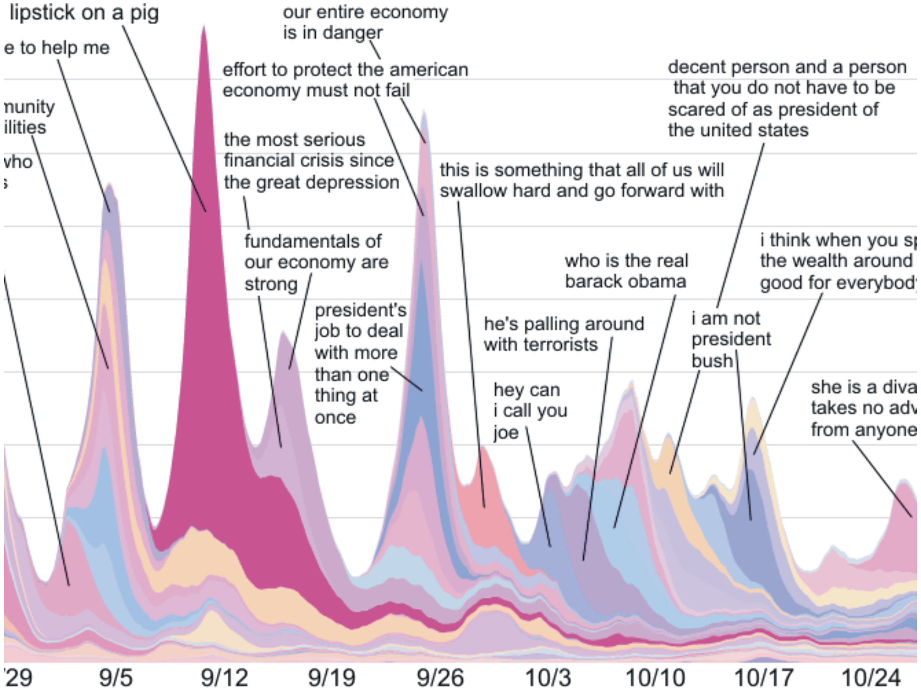
Application: Influential Blog Discovery[1]

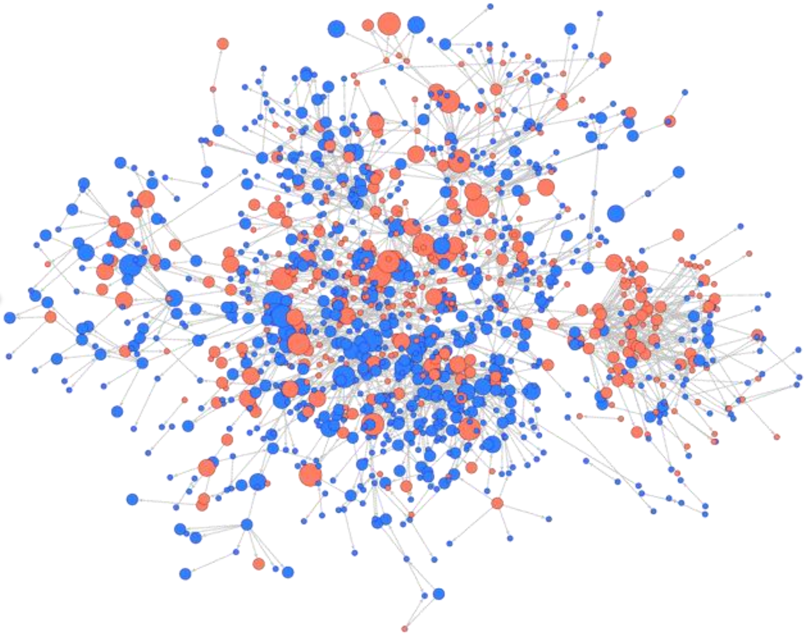
- Influential Blog Discovery
  - In the web 2.0 era, people spend a significant amount of time on user-generated content web sites, like blog sites.
  - Opinion leaders bring in new information, ideas, and opinions, and disseminate them down to the masses.
- Four properties for each bloggers
  - **Recognition**: A lot of inlinks to the article.
  - **Activity generation**: A large number of comments indicates that the blog is influential.
  - **Novelty**: with less outgoing links.
  - **Eloquence**: Longer articles tend to be more eloquent, and can thus be more influential.

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In WSDM'08, pages 207–217, 2008.

Submodular functions and their applications

Network Inference



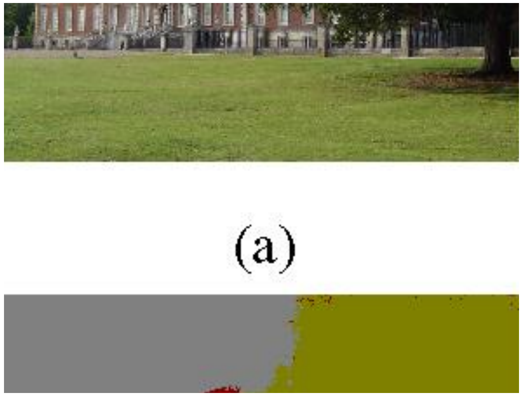


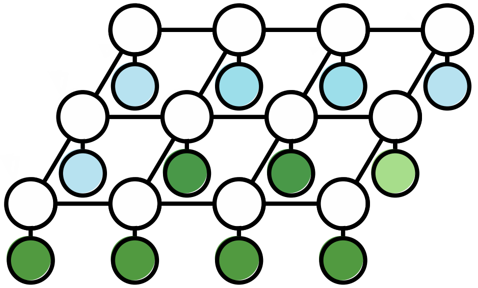
How learn who influences whom?

Summarizing Documents

How select representative sentences?

MAP (Maximum A-Posteriori) inference





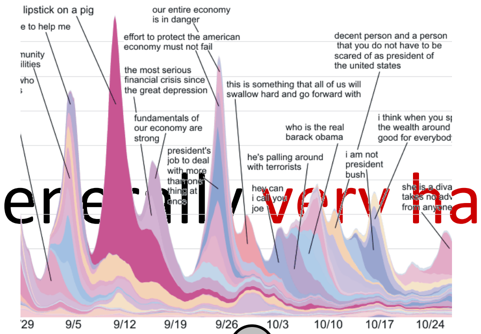
$$\max_{x} \quad p(x \mid z)$$

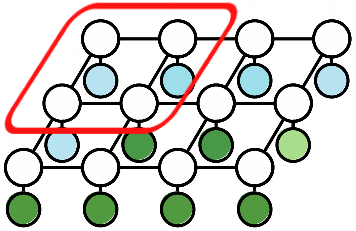
How find the MAP labeling in discrete graphical models *efficiently*?

What’s common?

- Formalization:

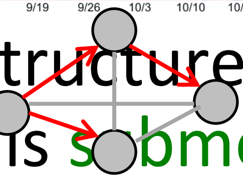
Optimize a set function F(S) under constraints





- generally very hard

- but: structure helps!
… if F is submodular, we can …



  - solve optimization problems with strong guarantees
  - solve some learning problems

Outline

- What is submodularity?
- Optimization
  - Minimization

many new results! ☺

Part I

  - Maximization
- Learning
- Learning for Optimization: new settings

Part II

Outline

many new results! ☺

- What is submodularity?
- Optimization
  - Minimization: new algorithms, constraints

Part I

  - Maximization: new algorithms (unconstrained)
- Learning
- Learning for Optimization: new settings

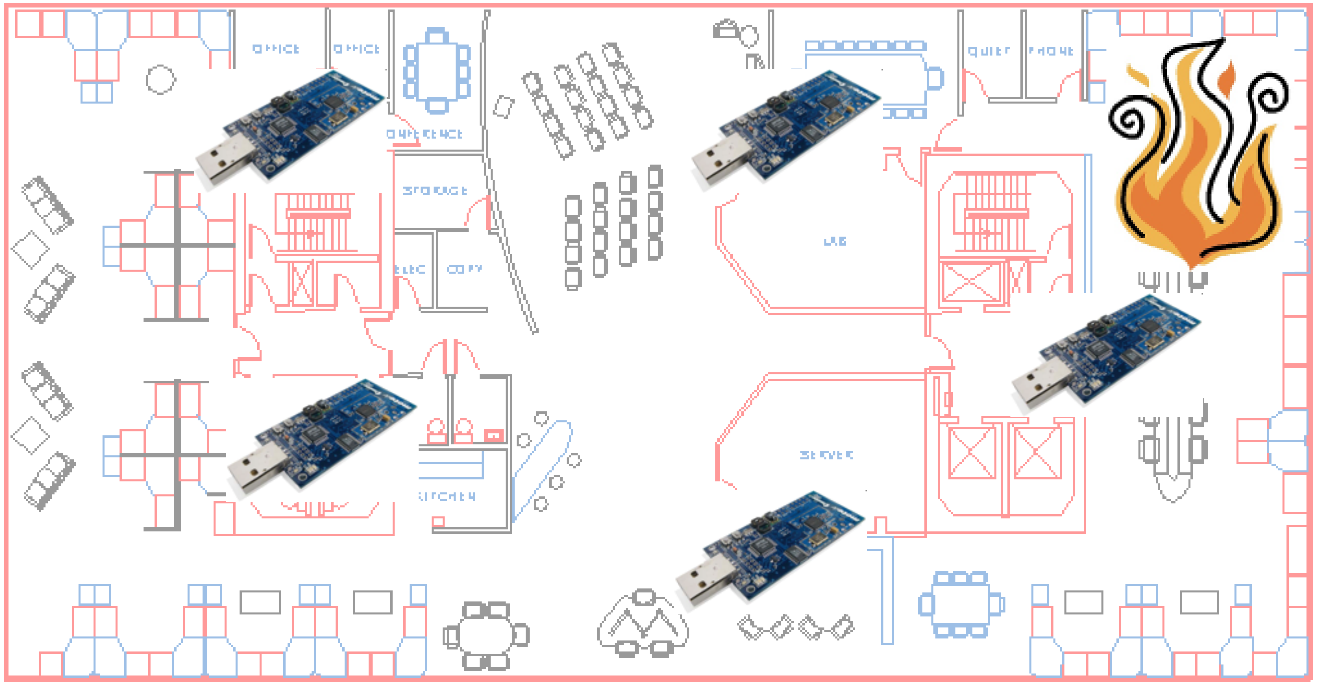
Part II

… and many new applications!

submodularity.org

slides, links, references, workshops, …

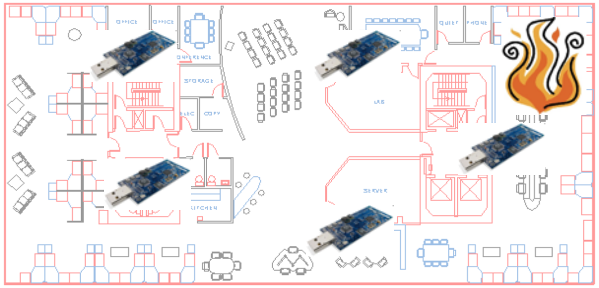
Example: placing sensors



Place sensors to monitor temperature

Set functions

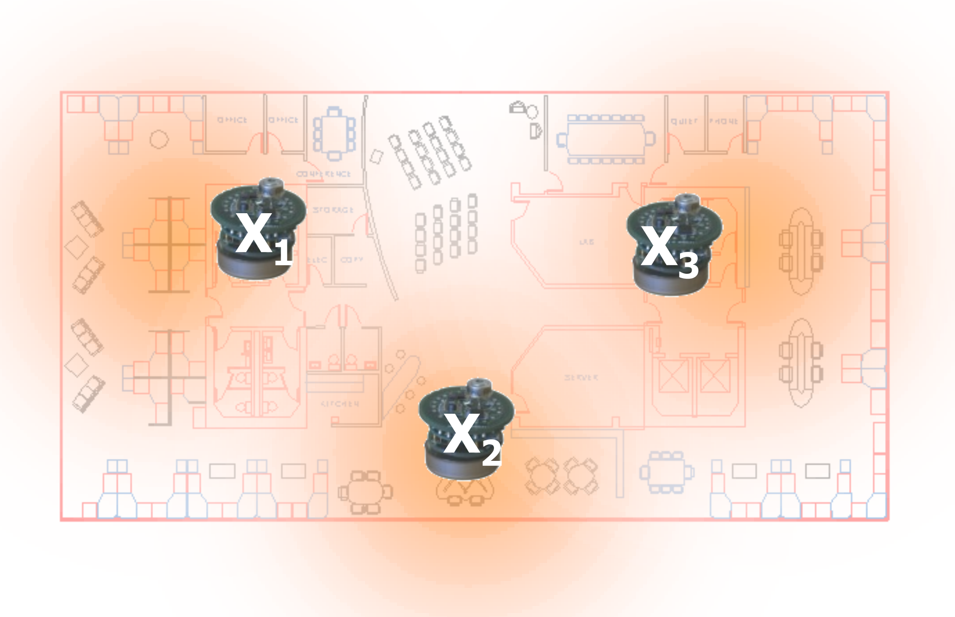
- finite ground set $V = \{1, 2, \ldots, n\}$
- set function $F : 2^V \to \mathbb{R}$



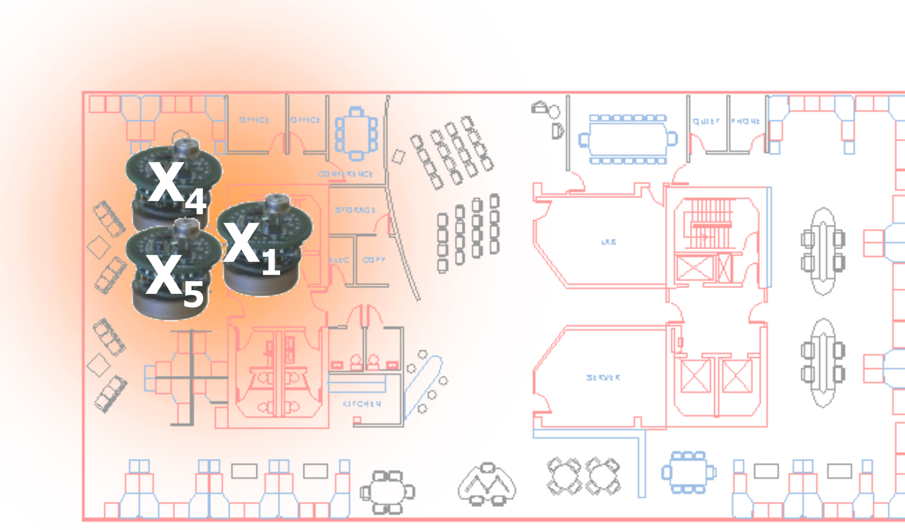
- will assume $F(\emptyset) = 0$ (w.l.o.g.)
- assume black box that can evaluate $F(A)$ for any $A \subseteq V$

Example: placing sensors

Utility $F(A)$ of having sensors at subset A of all locations



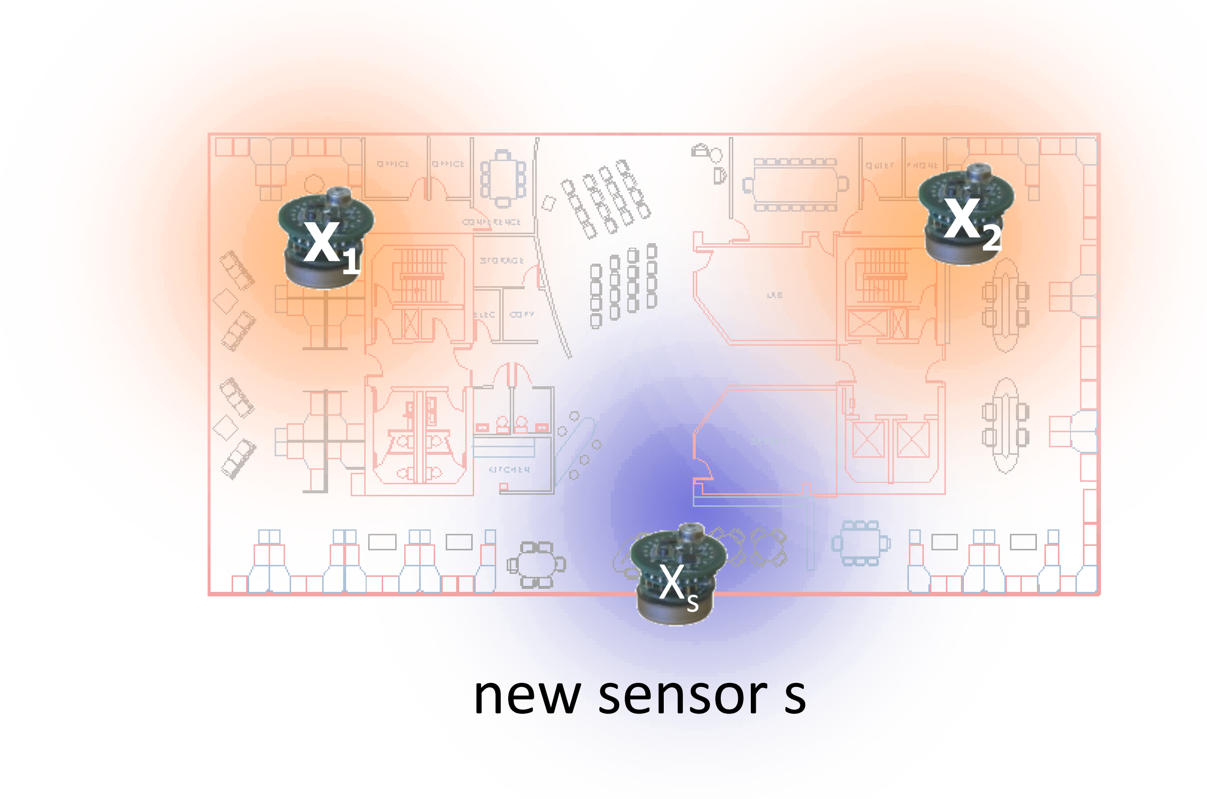
A={1,2,3}: Very informative
High value F(A)



A={1,4,5}: Redundant info
Low value F(A)

Marginal gain

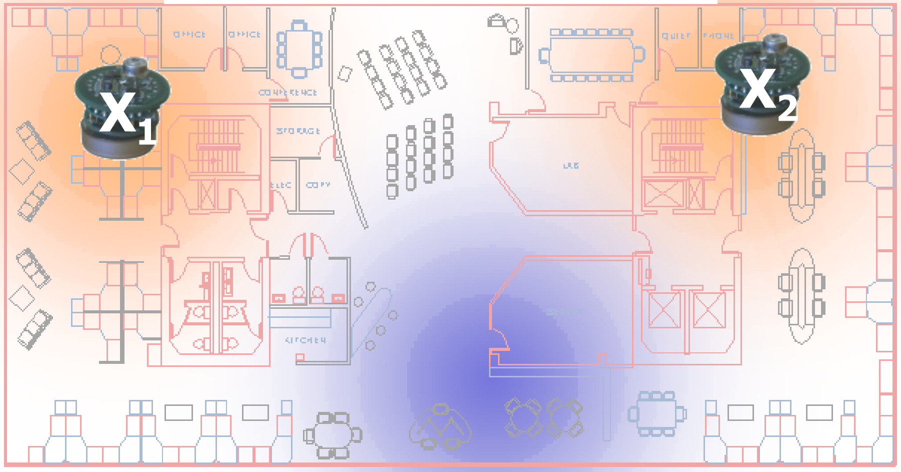
- Given set function $F : 2^V \rightarrow \mathbb{R}$
- Marginal gain: $\Delta_F(s \mid A) = F(\{s\} \cup A) - F(A)$



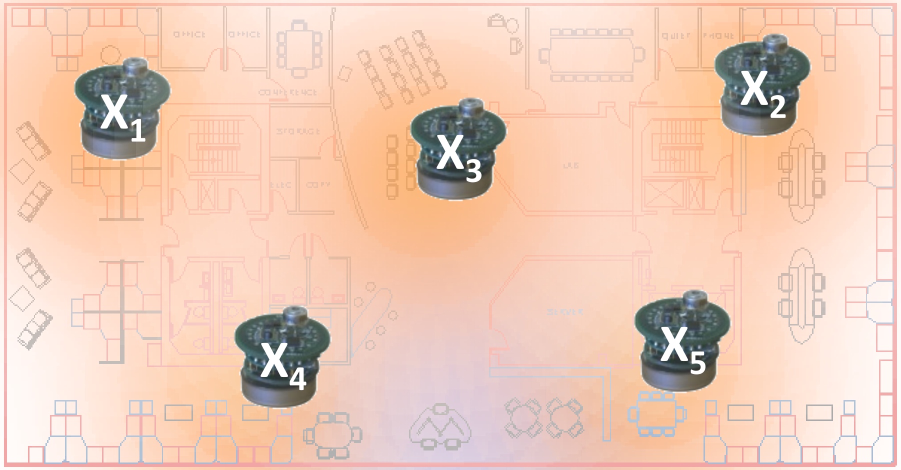
new sensor s

Decreasing gains: submodularity

placement A = {1,2}



placement B = {1,…,5}

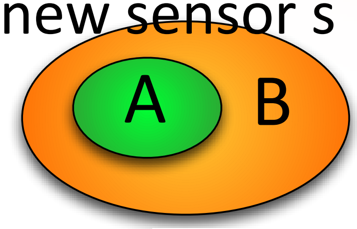


Big gain

new sensor s

small gain

+ • s



+ • s

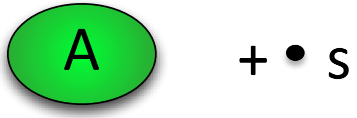
$$A \subseteq B$$

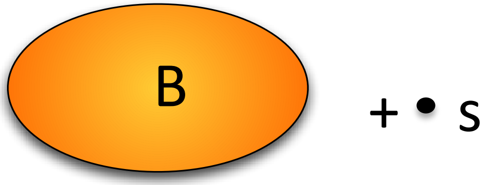
$$F(A \cup s) - F(A)$$

$$\Delta(s \mid A)$$

Equivalent characterizations

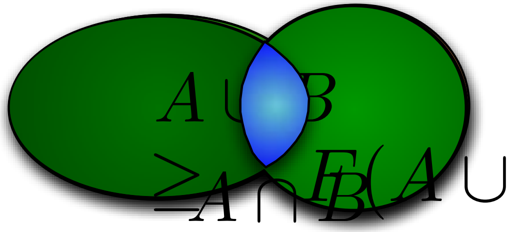
- Diminishing gains: for all $A \subseteq B$





$$F(A \cup s) - F(A) \quad \geq \quad F(B \cup s) - F(B)$$

- Union-Intersection: for all $A, B \subseteq V$



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

Questions

How do I prove my problem is submodular?

Why is submodularity useful?

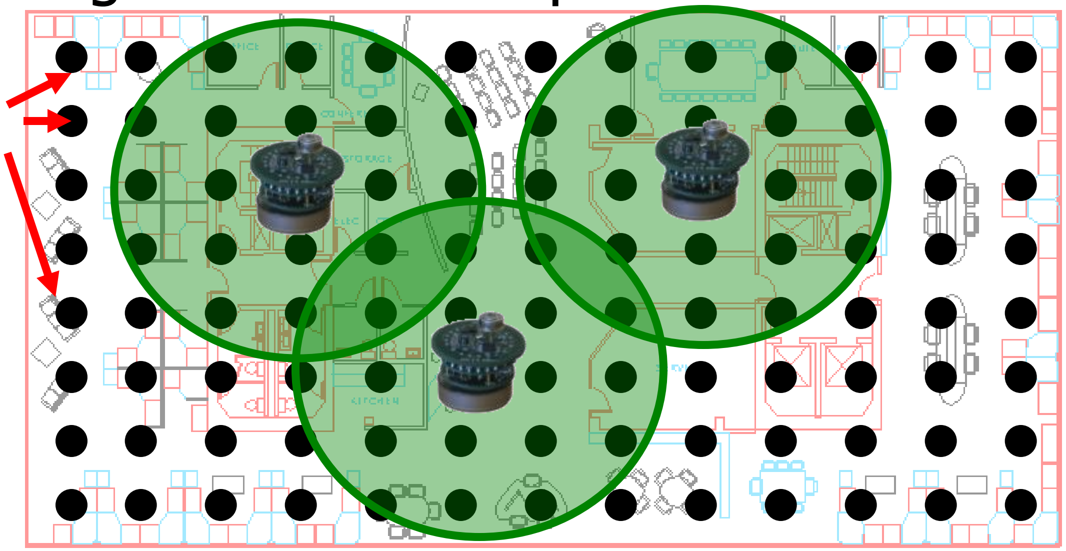
Example: Set cover

place sensors in building

Node predicts values of positions with some radius

goal: cover floorplan with discs

Possible locations V



$A \subseteq V$: $F(A) =$

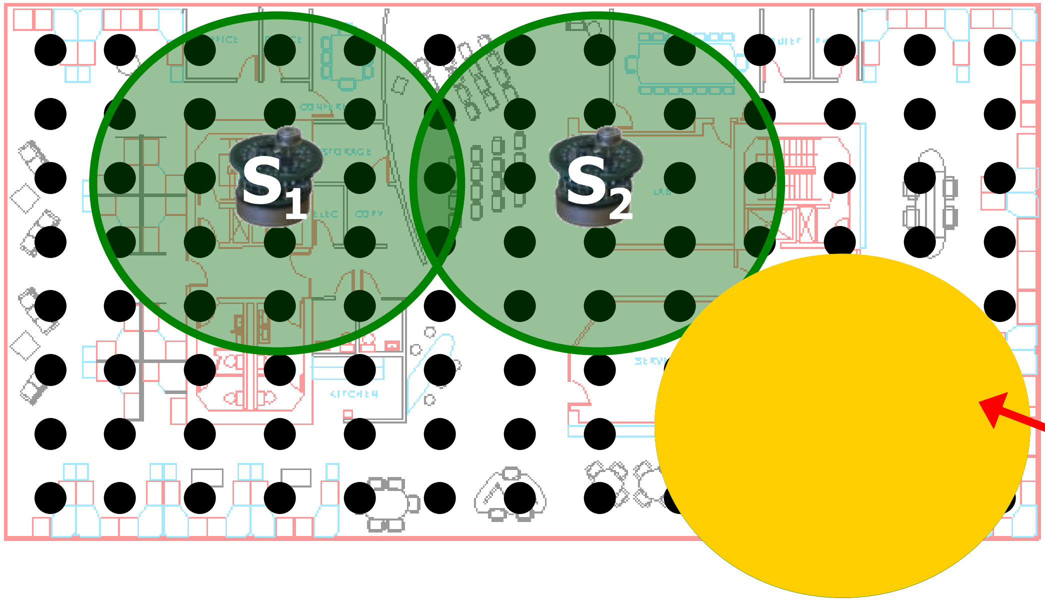
“area covered by sensors placed at A”

Formally:

Finite set W, collection of n subsets $S_i \subseteq W$

For $A \subseteq V$ define $F(A) = \left|\bigcup_{i \in A} S_i\right|$

Set cover is submodular

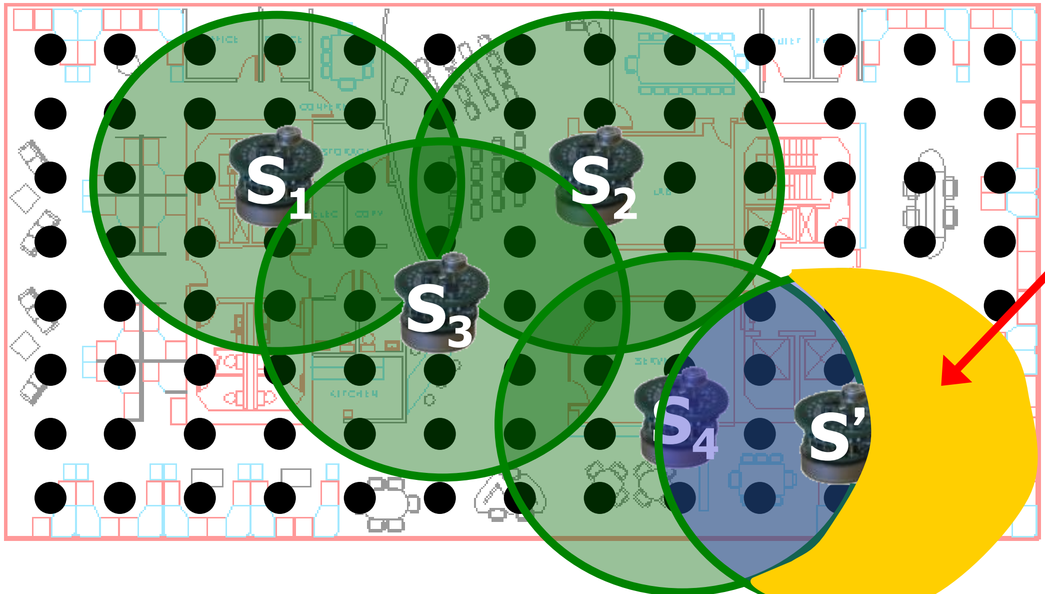


$A=\{s_1,s_2\}$

$F(A \cup \{s'\}) - F(A)$

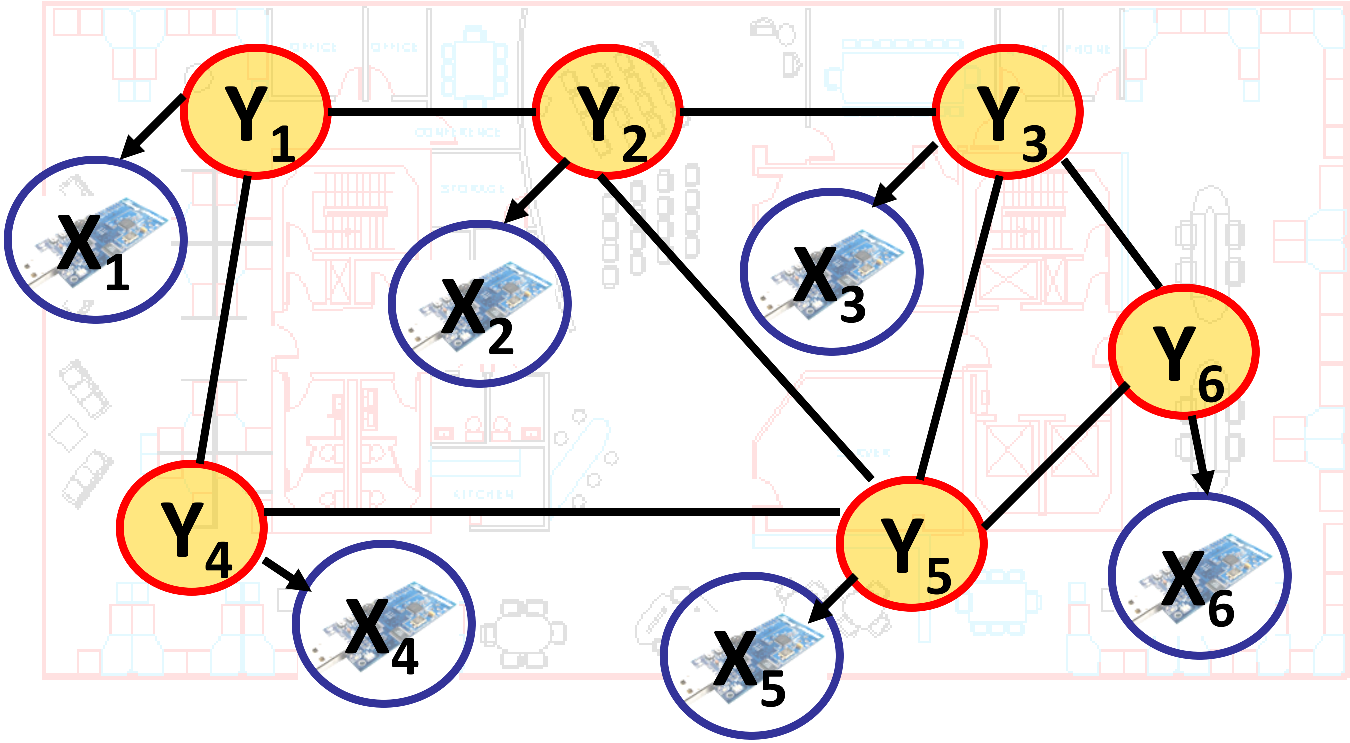
$\geq$

$F(B \cup \{s'\}) - F(B)$



$B = \{s_1,s_2,s_3,s_4\}$

More complex model for sensing



Y_s: temperature at location s

X_s: sensor value at location s

$X_s = Y_s$ + noise

Joint probability distribution

$$P(X_1,\ldots,X_n,Y_1,\ldots,Y_n) = \underbrace{P(Y_1,\ldots,Y_n)}_{\textbf{Prior}}\ \underbrace{P(X_1,\ldots,X_n \mid Y_1,\ldots,Y_n)}_{\textbf{Likelihood}}$$

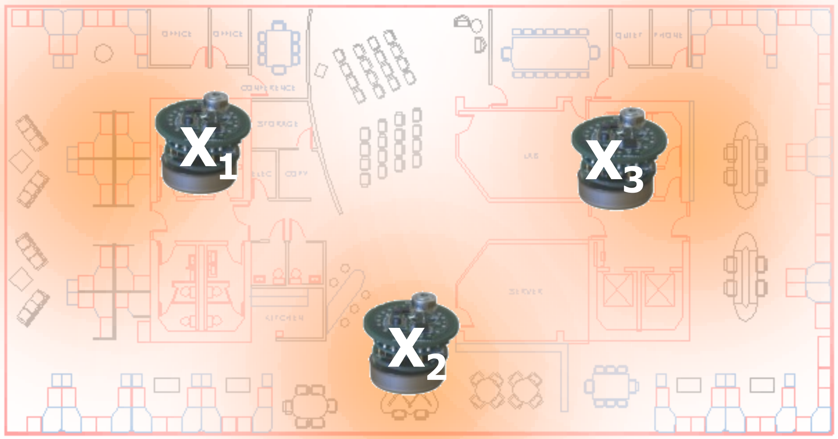
Example: Sensor placement

Utility of having sensors at subset A of all locations

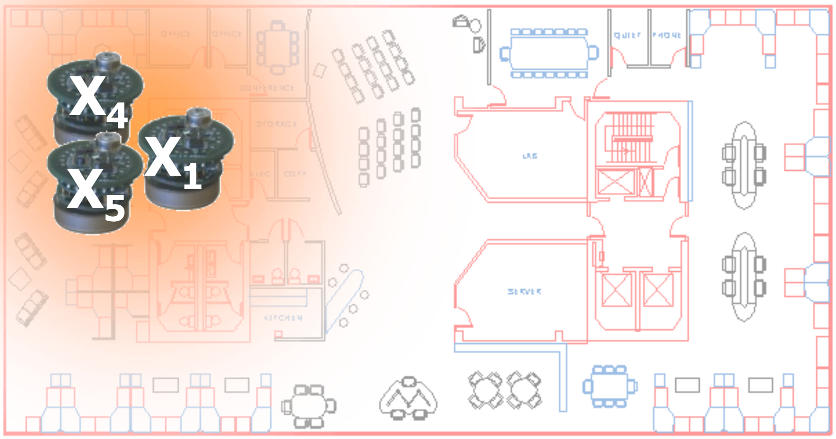
$$F(A) = H(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}_A)$$

Uncertainty
about temperature Y
before sensing

Uncertainty
about temperature Y
after sensing



A={1,2,3}: High value F(A)



A={1,4,5}: Low value F(A)

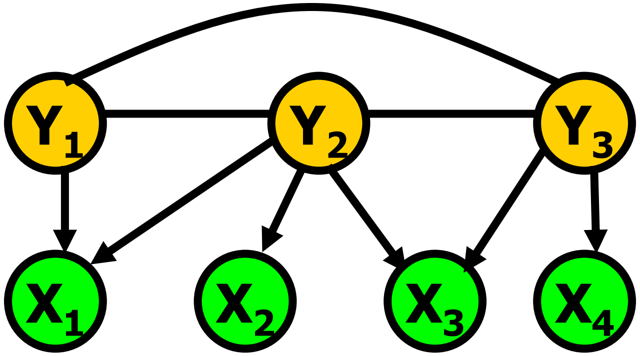
Submodularity of Information Gain

$Y_1,\ldots,Y_m, X_1, \ldots, X_n$ discrete RVs

$F(A) = I(Y; X_A) = H(Y)-H(Y \mid X_A)$

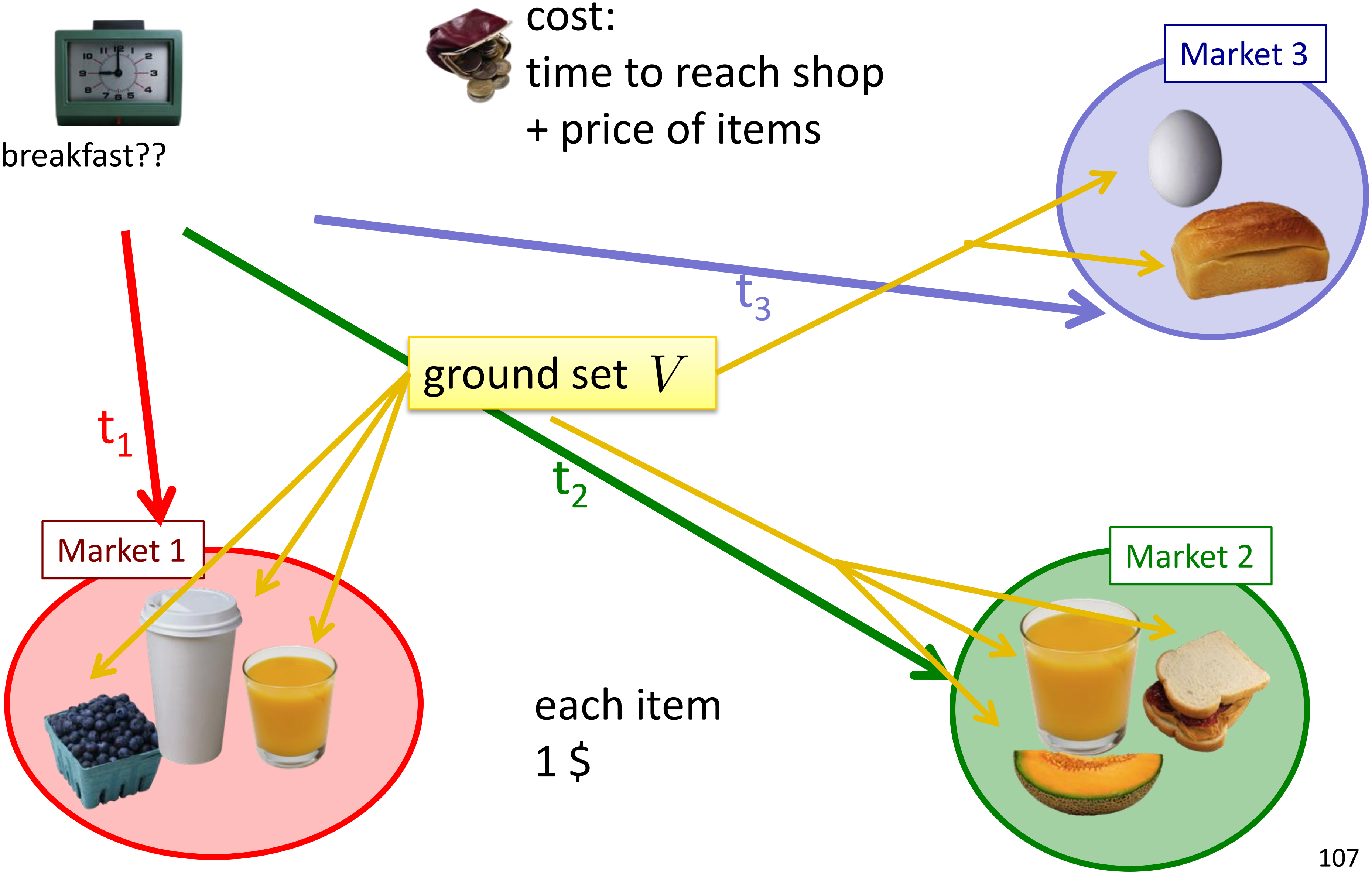
- F(A) is NOT always submodular

If X_i are all conditionally independent given Y, then F(A) is submodular! [Krause & Guestrin `05]



Proof:
"information never hurts"

Example: costs



breakfast??

cost:
time to reach shop
+ price of items

Market 3

t_3

ground set V

t_1

t_2

Market 1

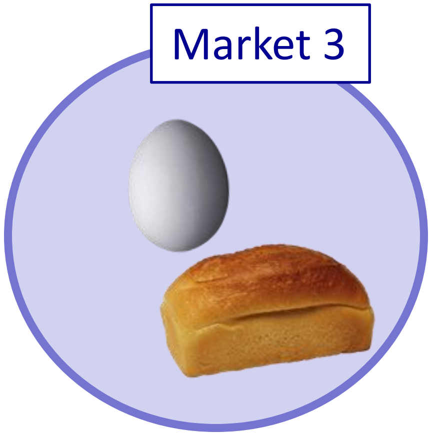
Market 2

each item
1 $

Example: costs

breakfast??

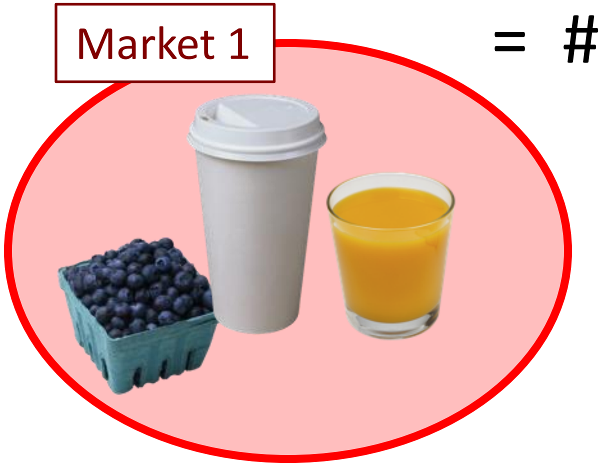
cost:
time to shop
+ price of items



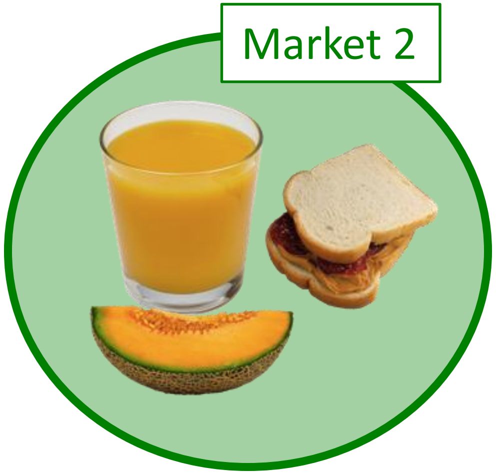
F( ) = cost( ) + cost(,)

$= t_1 + 1 + t_2 + 2$

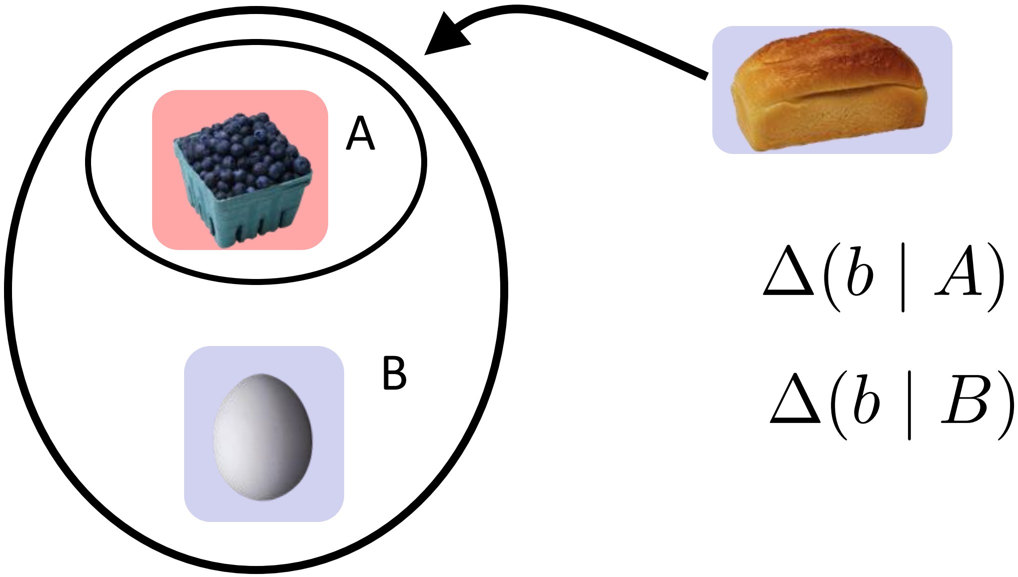
= #shops + #items



submodular?

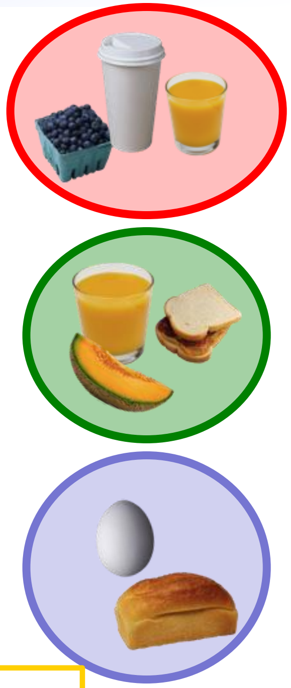


Shared fixed costs



$$\Delta(b \mid A) = 1 + t_3$$

$$\Delta(b \mid B) = 1$$

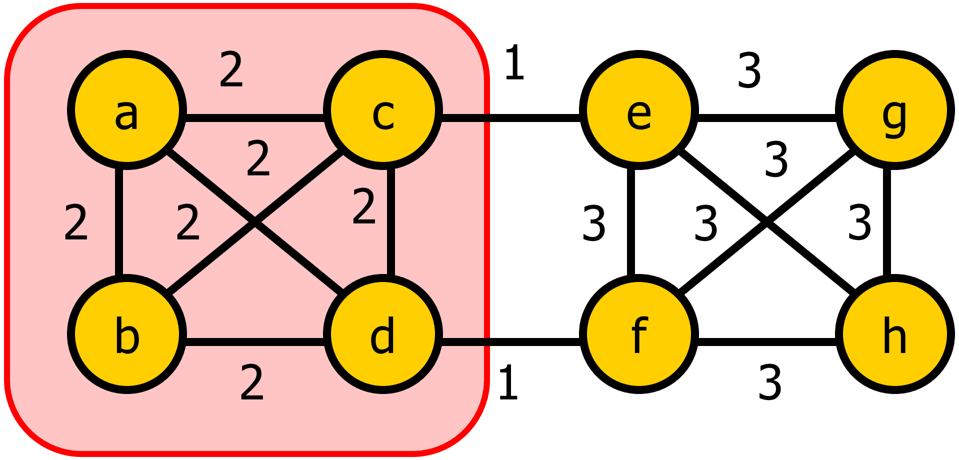


marginal cost: #new shops + #new items

decreasing → cost is submodular!

- shops: shared fixed cost
- economies of scale

Another example: Cut functions

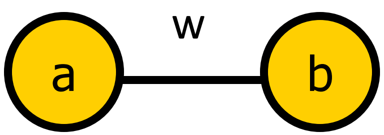


V={a,b,c,d,e,f,g,h}

$$F(A) = \sum_{s \in A, t \notin A} w_{s,t}$$

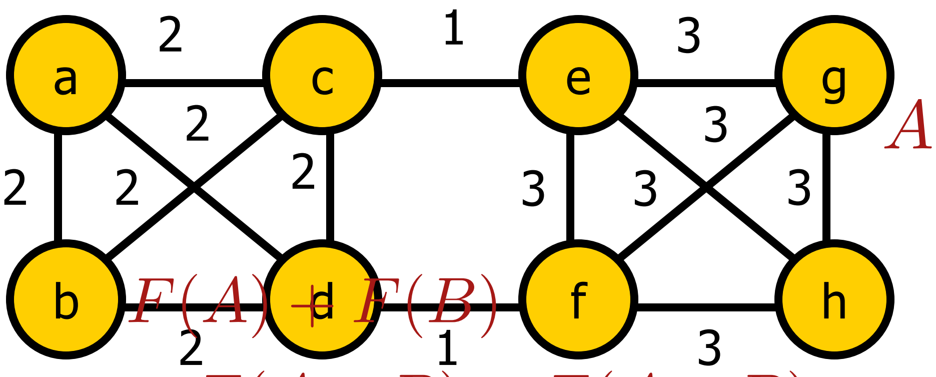
Cut function is submodular!

Why are cut functions submodular?



	S	F_{ab}(S)
$A \cap B$	{}	0
A	{a}	w
B	{b}	w
$A \cup B$	{a,b}	0

Submodular if w ≥ 0!



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

$$F(S) = \sum_{(i,j)\in E} \underbrace{F_{i,j}(S \cap \{i,j\})}$$

Cut function in subgraph {i,j}

➔ Submodular!

Closedness properties

$F_1,\ldots,F_m$ submodular functions on V and $\lambda_1,\ldots,\lambda_m > 0$

Then: $F(A) = \sum_i \lambda_i F_i(A)$ is submodular

Submodularity closed under nonnegative linear combinations!

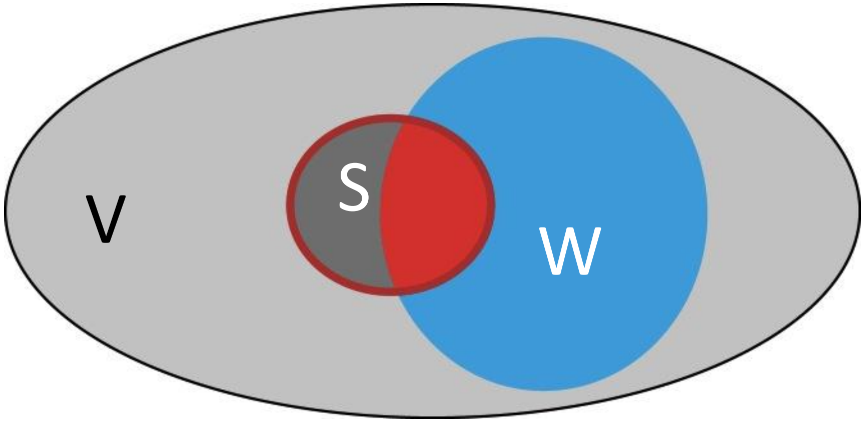
Extremely useful fact:

- $F_\theta(A)$ submodular ➔ $\sum_\theta P(\theta) F_\theta(A)$ submodular!
- Multicriterion optimization
- A basic proof technique! ☺

Other closedness properties

- Restriction: F(S) submodular on V, W subset of V

  Then $F'(S) = F(S \cap W)$ is submodular



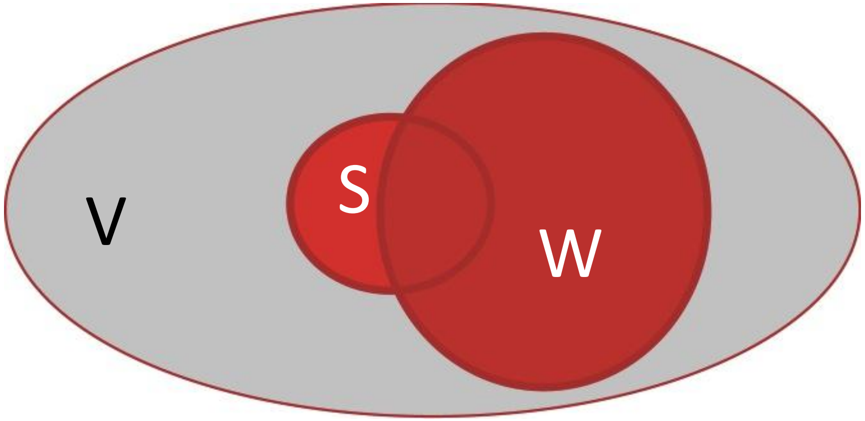
Other closedness properties

- Restriction: F(S) submodular on V, W subset of V

  Then $F'(S) = F(S \cap W)$ is submodular

- Conditioning: F(S) submodular on V, W subset of V

  Then $F'(S) = F(S \cup W)$ is submodular



Other closedness properties

- Restriction: F(S) submodular on V, W subset of V

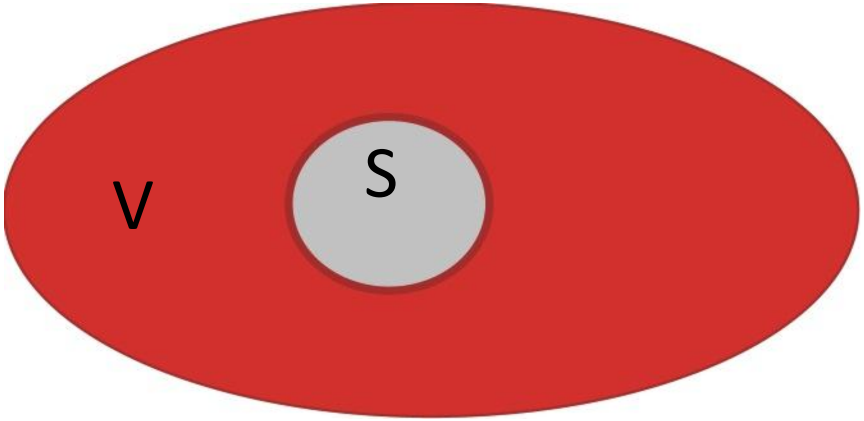
  Then $F'(S) = F(S \cap W)$ is submodular

- Conditioning: F(S) submodular on V, W subset of V

  Then $F'(S) = F(S \cup W)$ is submodular

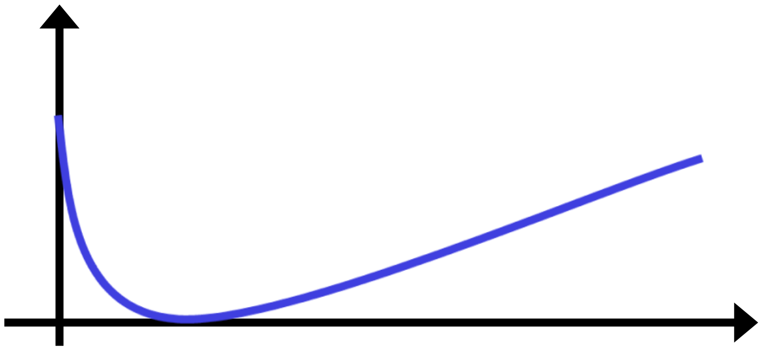
- Reflection: F(S) submodular on V

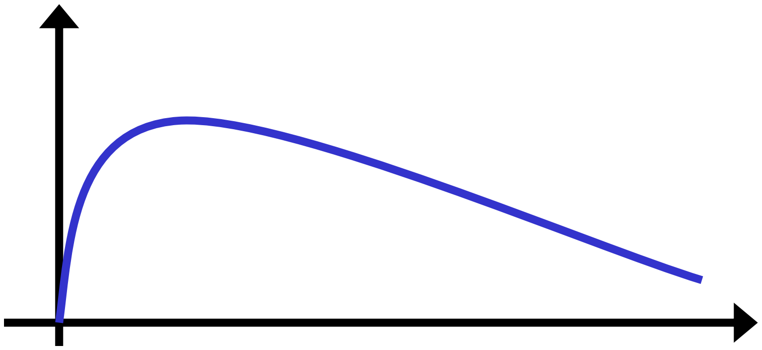
  Then $F'(S) = F(V \setminus S)$ is submodular



Submodularity …

discrete convexity ….

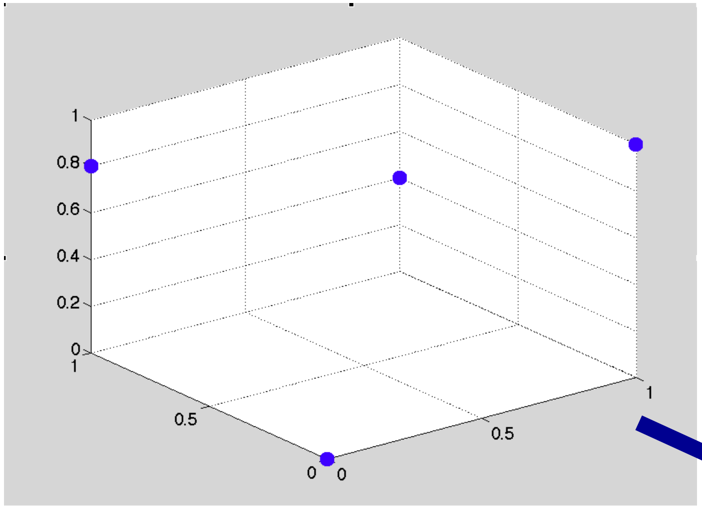


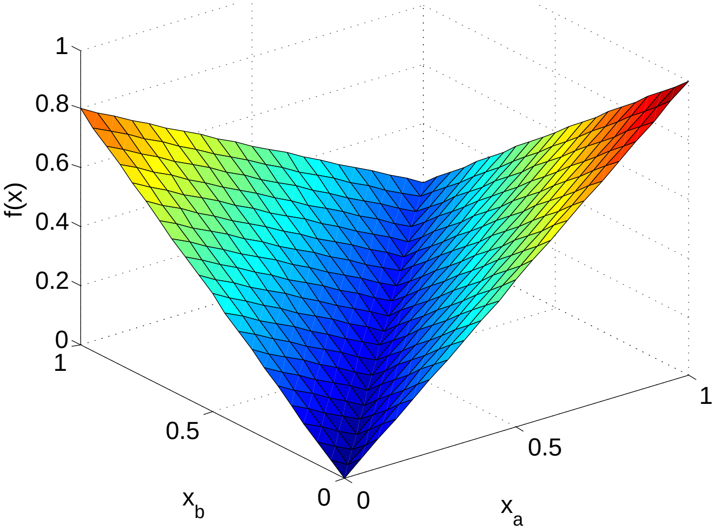


… or concavity?

Convex aspects

- convex extension
  - duality
  - efficient minimization





But this is only half of the story…

Concave aspects

- submodularity:

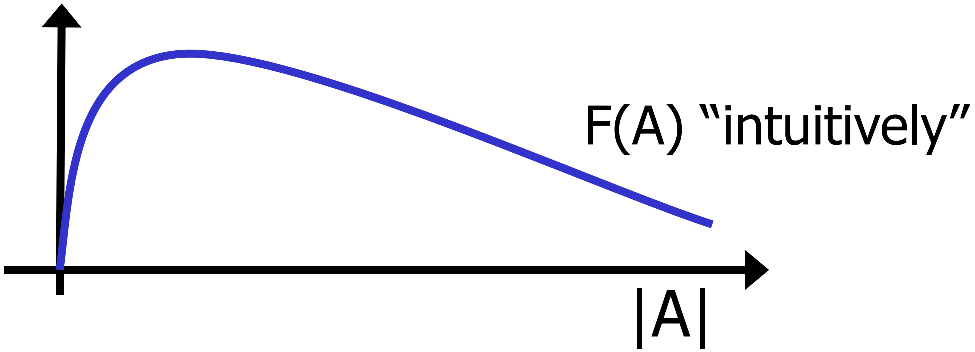
$$A \subseteq B, \quad s \notin B :$$

$$F(A \cup s) - F(A) \quad \geq \quad F(B \cup s) - F(B)$$

- concavity:

$$a \leq b, \quad s > 0 :$$

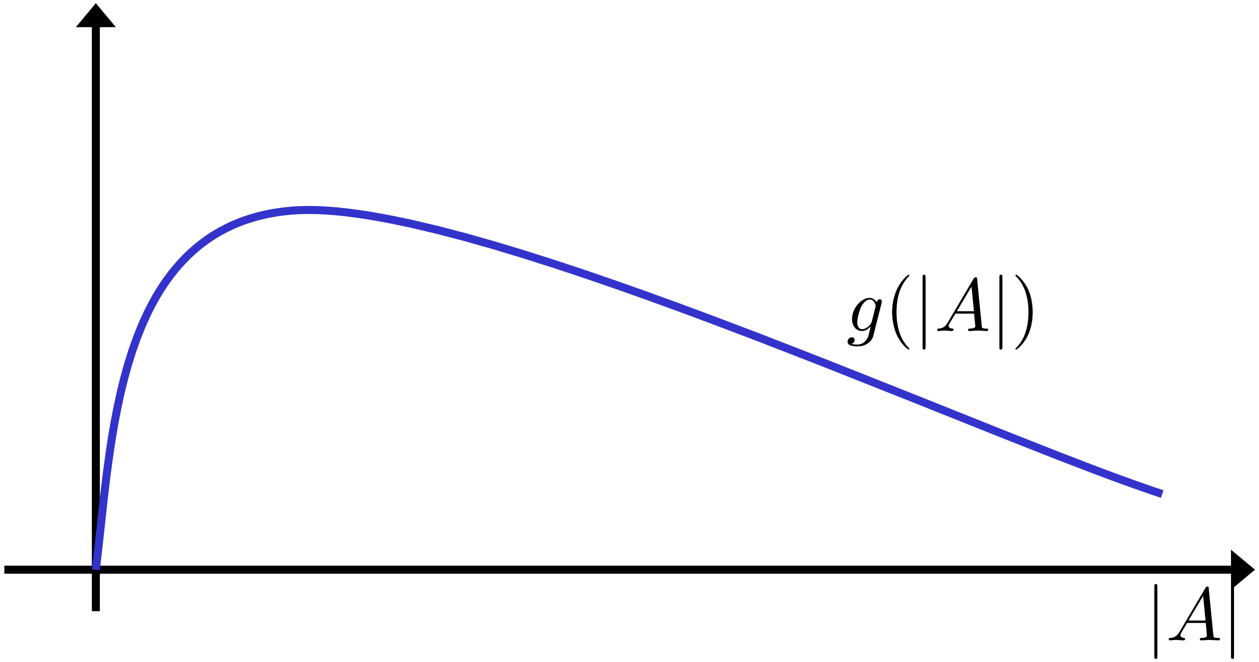
$$f(a+s) - f(a) \quad \geq \quad f(b+s) - f(b)$$



Submodularity and concavity

- suppose $g : \mathbb{N} \to \mathbb{R}$ and $F(A) = g(|A|)$

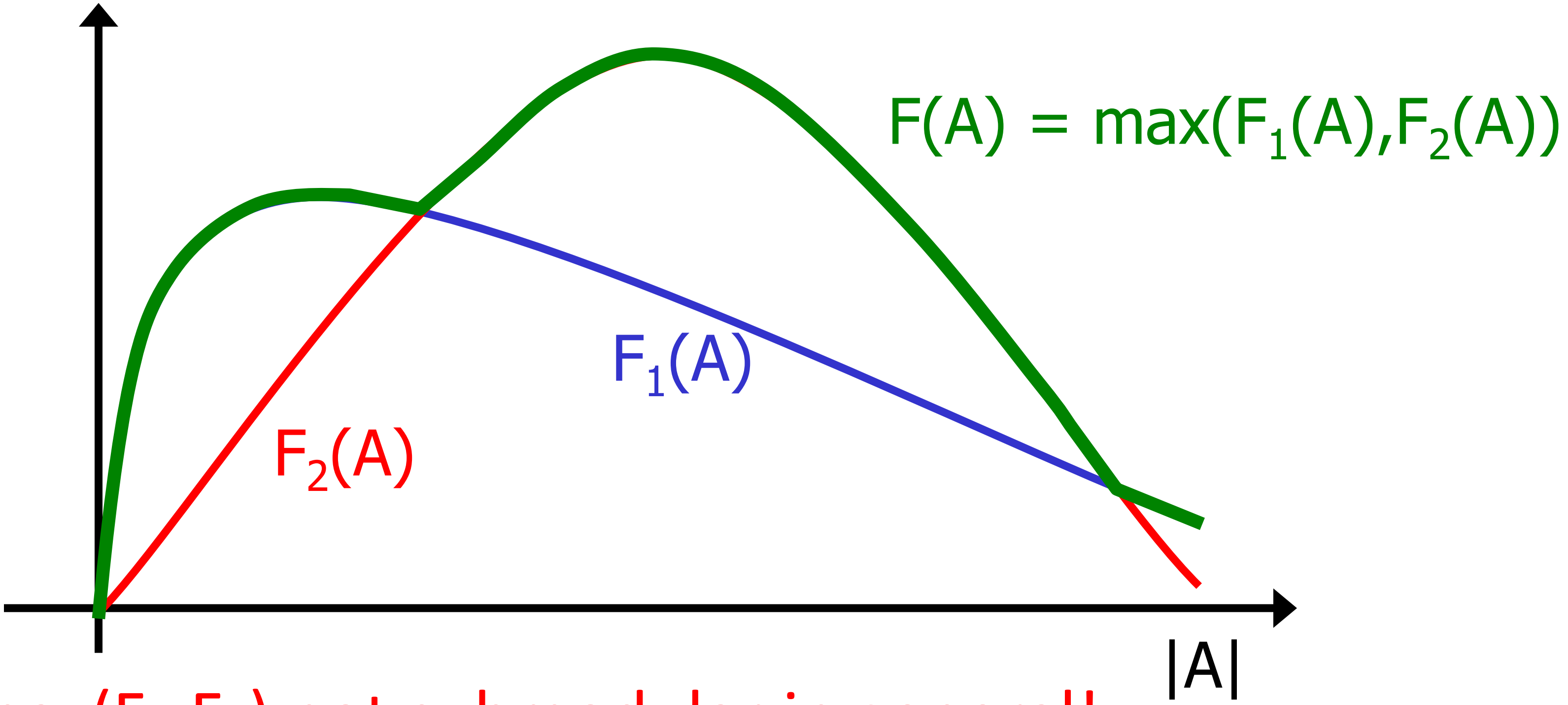
$F(A)$ submodular *if and only if* … g is concave



Maximum of submodular functions

- $F_1(A), F_2(A)$ submodular. What about

$$F(A) = \max\{\, F_1(A), F_2(A)\, \} \quad ?$$



max(F_1,F_2) not submodular in general!

Minimum of submodular functions

Well, maybe $F(A) = \min(F_1(A), F_2(A))$ instead?

	$F_1(A)$	$F_2(A)$
{}	0	0
{a}	1	0
{b}	0	1
{a,b}	1	1

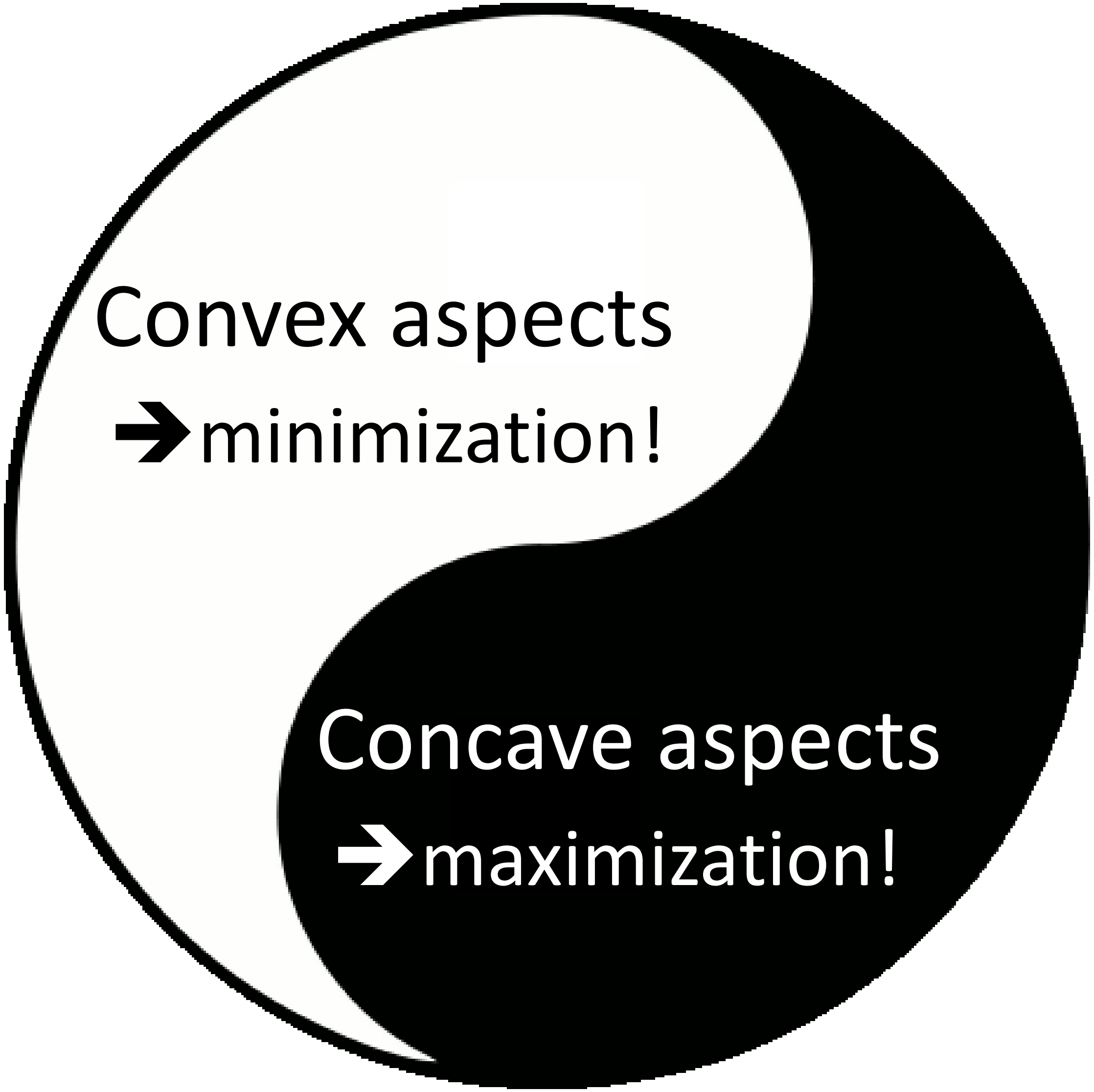
$$F(\{b\}) - F(\{\}) = 0$$

$$<$$

$$F(\{a,b\}) - F(\{a\}) = 1$$

$\min(F_1, F_2)$ not submodular in general!

Two faces of submodular functions



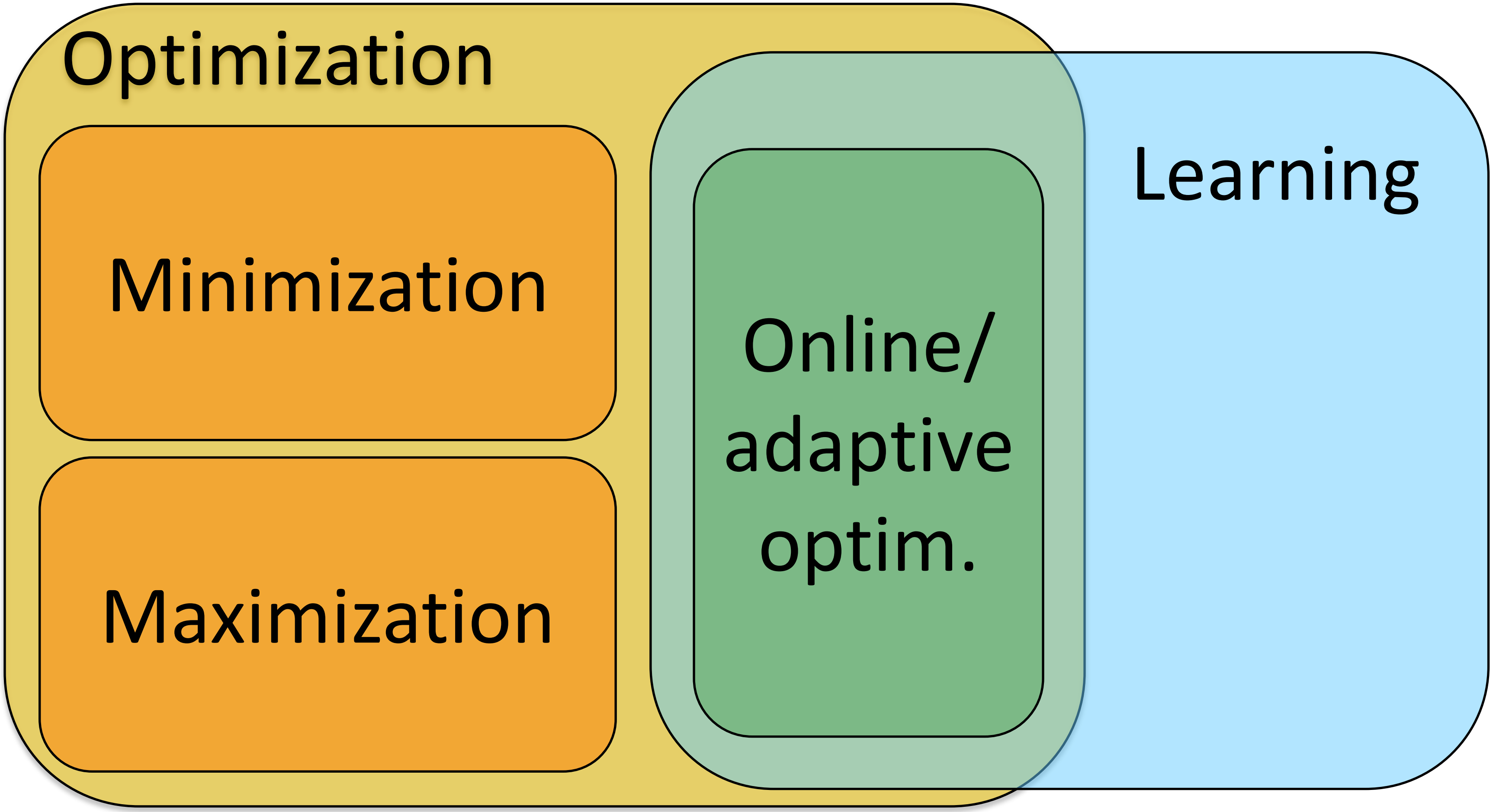
Convex aspects

→minimization!

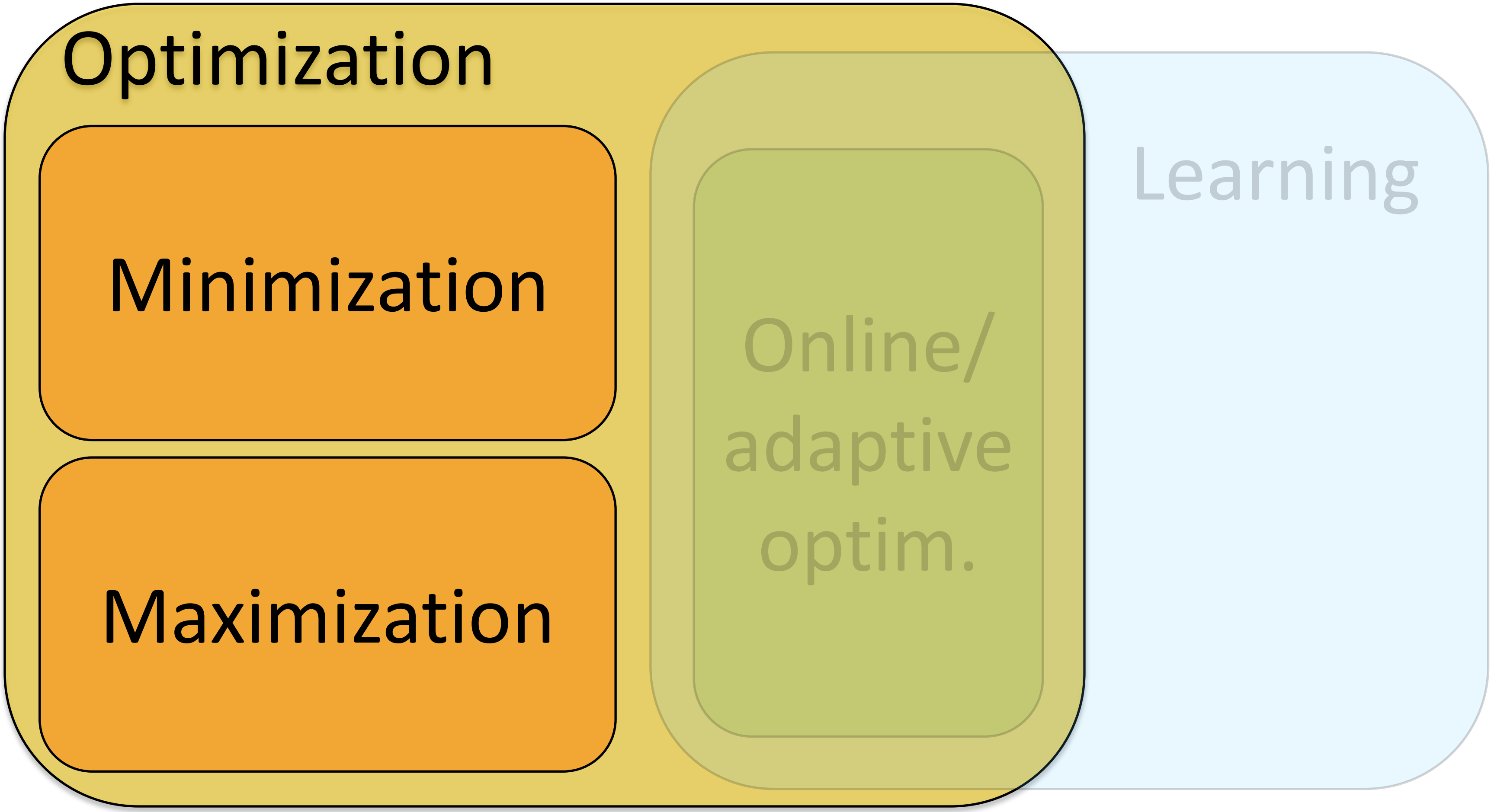
Concave aspects

→maximization!

What to do with submodular functions

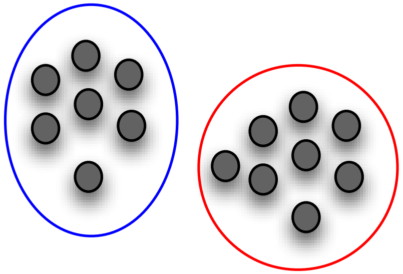


What to do with submodular functions

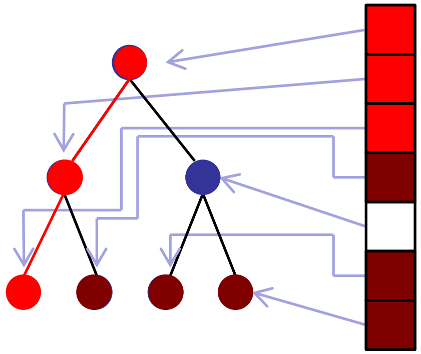


Minimization and maximization not the same??

Submodular minimization

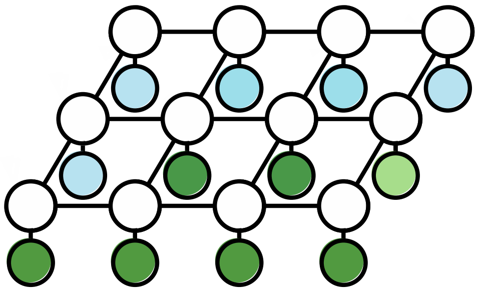


clustering

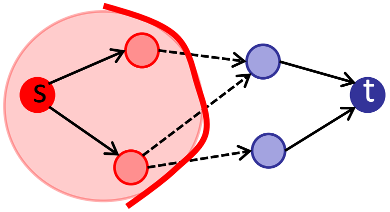


structured sparsity regularization

$$\min_{S\subseteq V} F(S)$$



MAP inference



minimum cut

Submodular minimization

$$\min_{S \subseteq V} F(S)$$

➔ submodularity and convexity

Set functions and energy functions

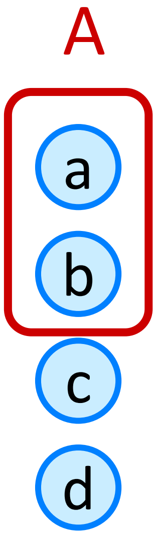
any set function

with $|V| = n$

$$F : 2^V \rightarrow \mathbb{R}$$

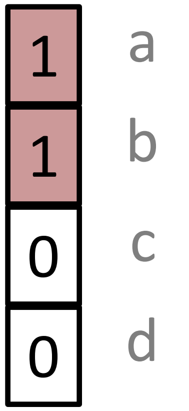
… is a function on binary vectors!

$$F : \{0,1\}^n \rightarrow \mathbb{R}$$



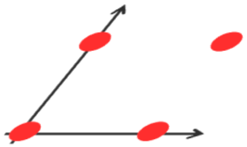
$$\hat{=}$$

$$x = e_A$$



pseudo-boolean function

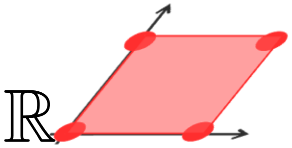
Submodularity and convexity



extension

$$f : [0,1]^n \to \mathbb{R}$$

$$F : \{0,1\}^n \to \mathbb{R}$$



Lovász extension

$$f(x) \;=\; \max_{y \in P_F} \; x \cdot y$$

convex

Lovász, 1982

- minimum of *f* is a minimum of F
- submodular minimization as convex minimization: polynomial time!

Grötschel, Lovász, Schrijver 1981

Submodularity and convexity

extension

$F : \{0,1\}^n \to \mathbb{R}$ → $f : [0,1]^n \to \mathbb{R}$

Lovász extension

$$f(x) = \max_{y \in P_F} x \cdot y$$

convex

Lovász, 1982

- minimum of *f* is a minimum of F
- submodular minimization as convex minimization: polynomial time!

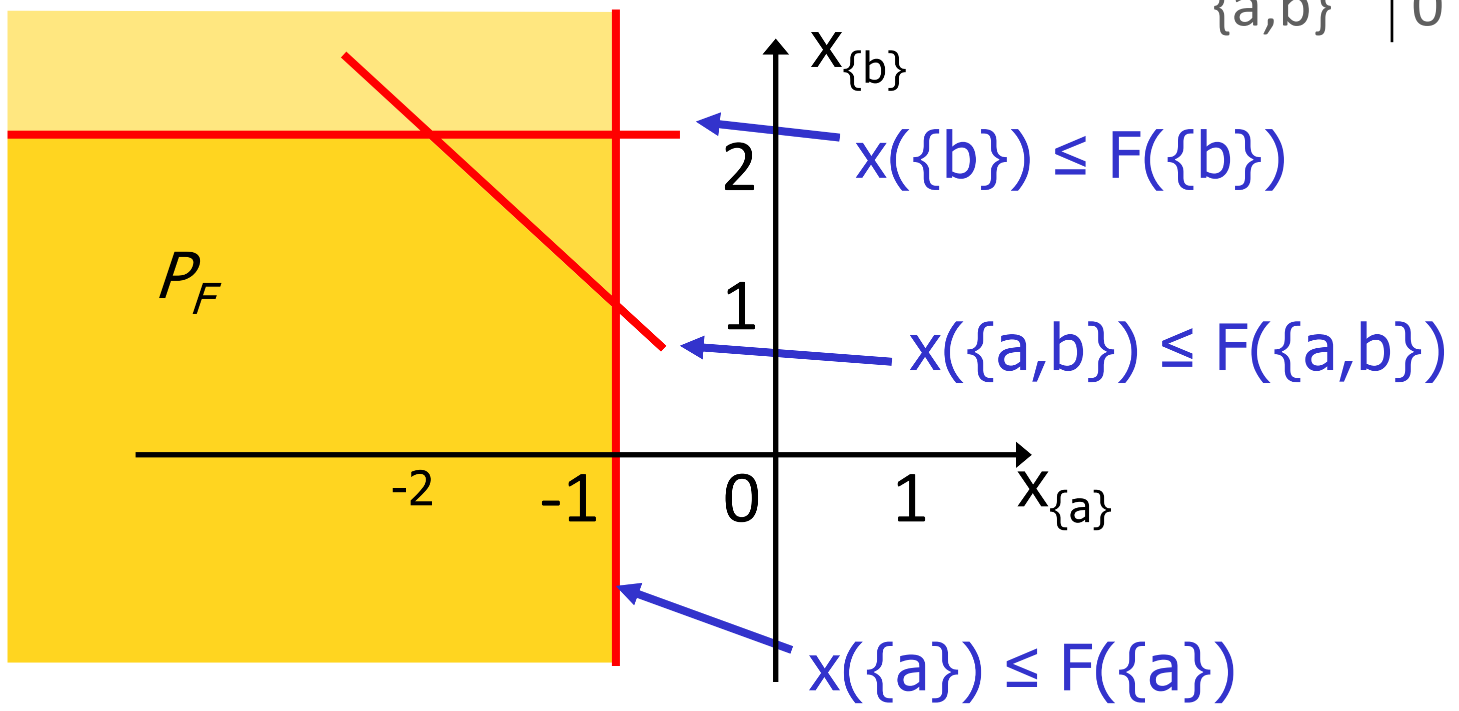
The submodular polyhedron P_F

$$P_F = \{x \in \mathbb{R}^n : x(A) \leq F(A) \text{ for all } A \subseteq V\}$$

$$x(A) = \sum_{i \in A} x_i$$

Example: V = {a,b}

A	F(A)
{}	0
{a}	-1
{b}	2
{a,b}	0



x({b}) ≤ F({b})

x({a,b}) ≤ F({a,b})

x({a}) ≤ F({a})

Evaluating the Lovász extension

$$P_F = \{x \in \mathbb{R}^n : x(A) \leq F(A) \text{ for all } A \subseteq V\}$$

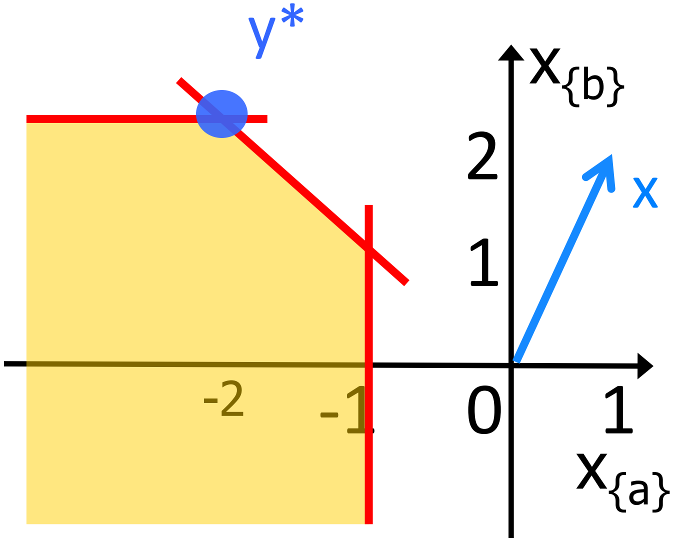
Linear maximization over P_F

$$f(x) = \max_{y \in P_F} x \cdot y$$

Exponentially many constraints!!! ☹

Computable in O(n log n) time ☺

[Edmonds '70]



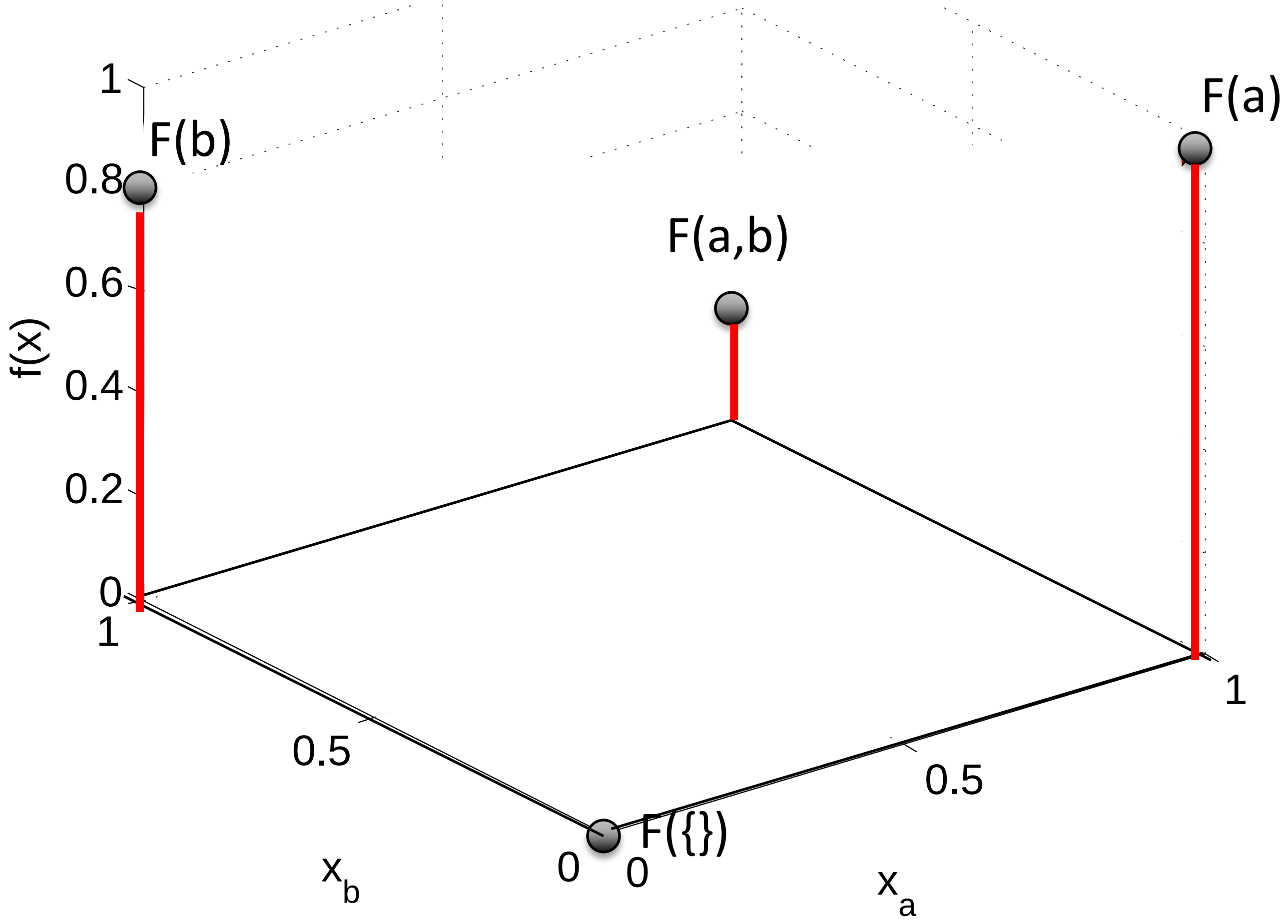
greedy algorithm:

- sort x
- order defines sets $S_i = \{1, \ldots, i\}$
- $y_i = F(S_i) - F(S_{i-1})$

➡

- Subgradient
- Separation oracle

Lovász extension: example



A	F(A)
{}	0
{a}	1
{b}	.8
{a,b}	.2

Submodular minimization

$$\min_{A \subseteq V} F(A)$$

minimize convex extension

- ellipsoid algorithm

  [Grötschel et al. `81]

- subgradient method, smoothing [Stobbe & Krause `10]

- duality: minimum norm point algorithm

  [Fujishige & Isotani '11]

combinatorial algorithms

- Fulkerson prize

  Iwata, Fujishige, Fleischer '01 & Schrijver '00

- state of the art:

  $O(n^4T + n^5 \log M)$ [Iwata '03]

  $O(n^6 + n^5T)$ [Orlin '09]

T = time for evaluating *F*

The minimum-norm-point algorithm

Example: V = {a,b}

Lovász extension

Regularized problem

A	F(A)
{}	0
{a}	-1
{b}	2
{a,b}	0

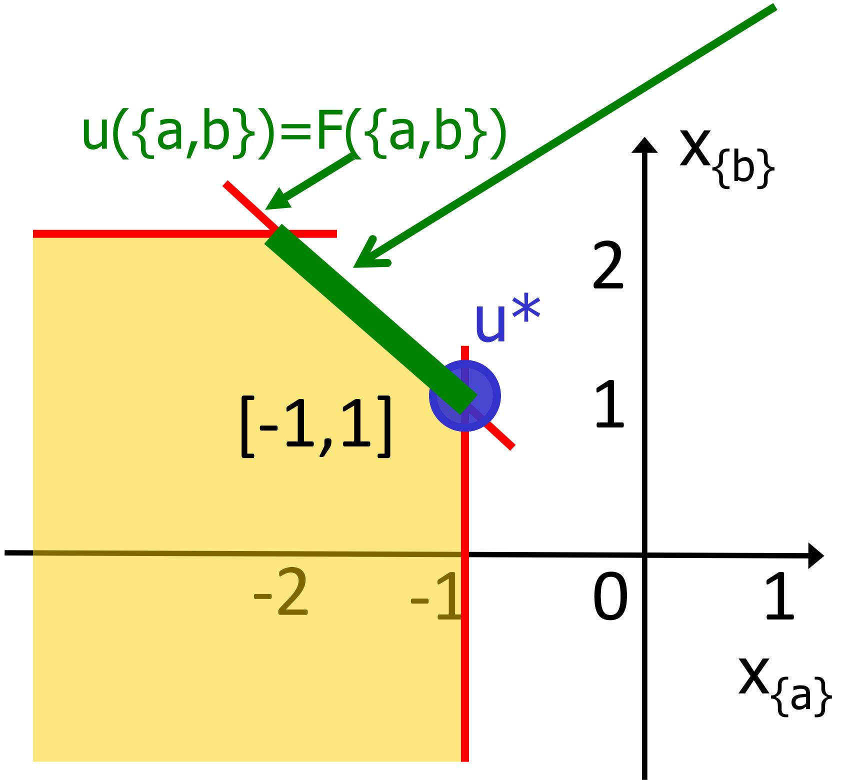
$$\min_{x\in[0,1]^n} f(x)$$

$$\min_x f(x) + \frac{1}{2}\|x\|^2$$

dual: minimum norm problem

$$u^* = \arg\min_{u\in B_F} \frac{1}{2}\|u\|^2$$

Base polytope B_F



$$A^* = \{i \mid u^*(i) \leq 0\}$$

minimizes *F*:

$$A^* = \arg\min_{A\subseteq V} F(A)$$

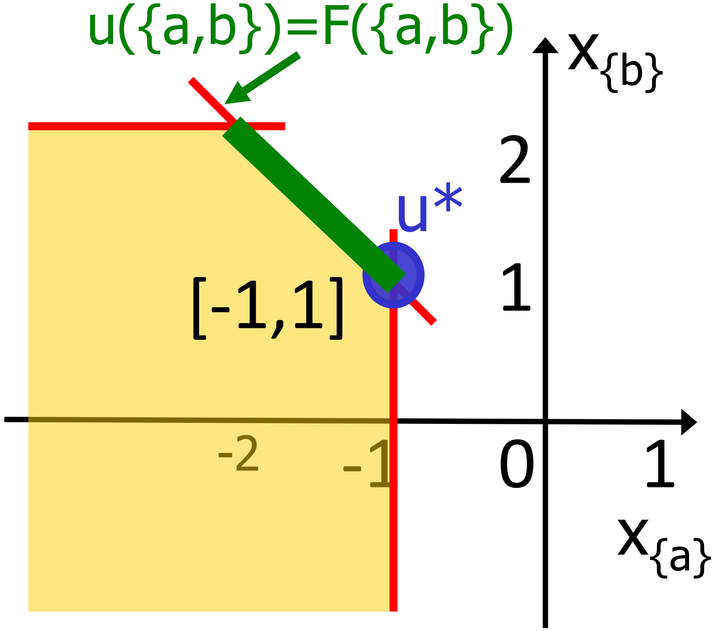
Fujishige ‘91, Fujishige & Isotani ‘11

The minimum-norm-point algorithm

1. find $u^* = \arg\min_{u \in B_F} \frac{1}{2}\|u\|^2$

2. $A^* = \{i \mid u^*(i) \leq 0\}$

can we solve this??



yes! ☺

recall: can solve

linear optimization over P_F

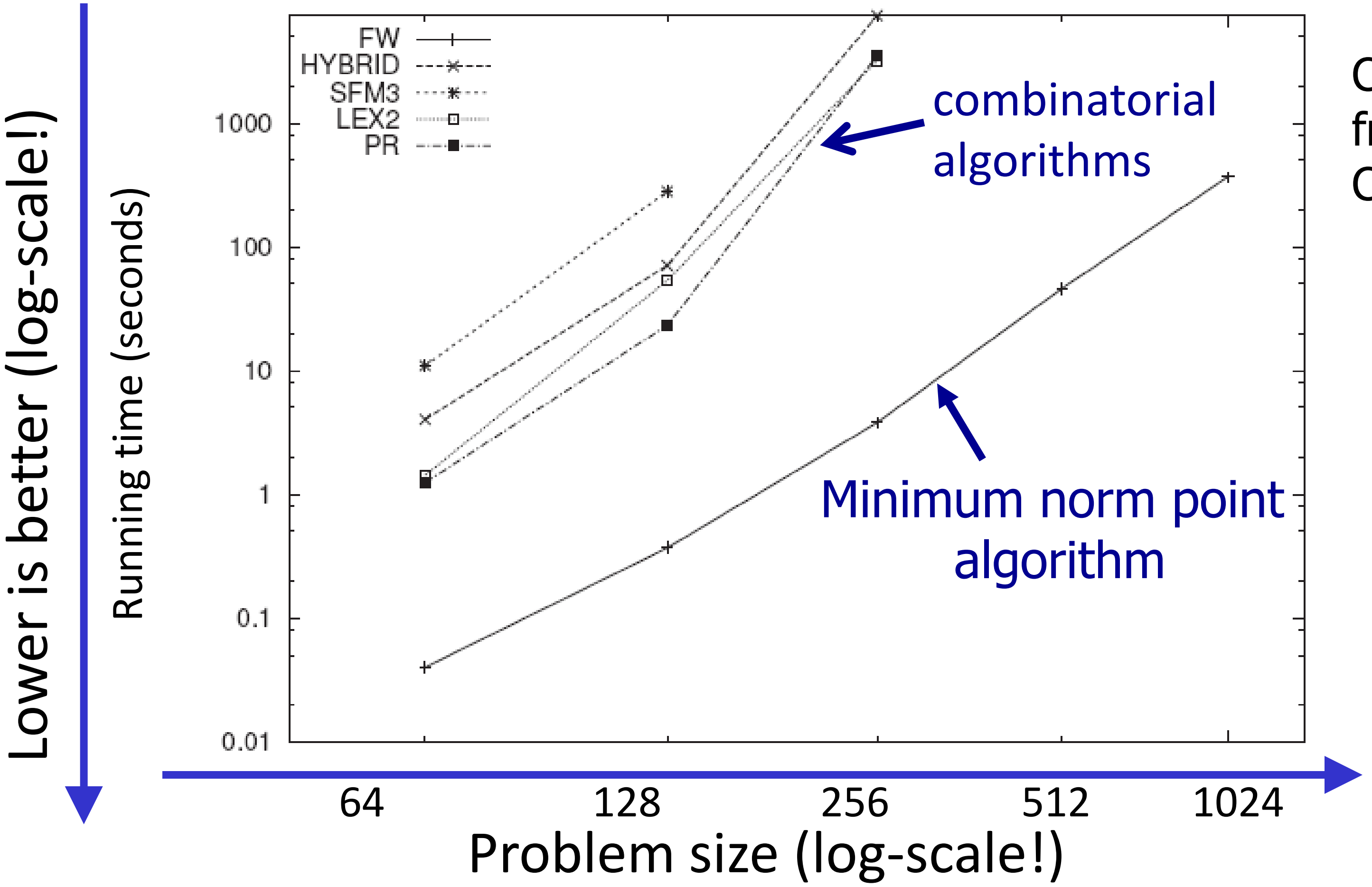
similar: optimization over B_F

➔ can find u^*

(Frank-Wolfe algorithm)

Fujishige '91, Fujishige & Isotani '11

Empirical comparison



Lower is better (log-scale!)

Running time (seconds)

combinatorial algorithms

Minimum norm point algorithm

64

128

256

512

1024

Problem size (log-scale!)

Cut functions from DIMACS Challenge

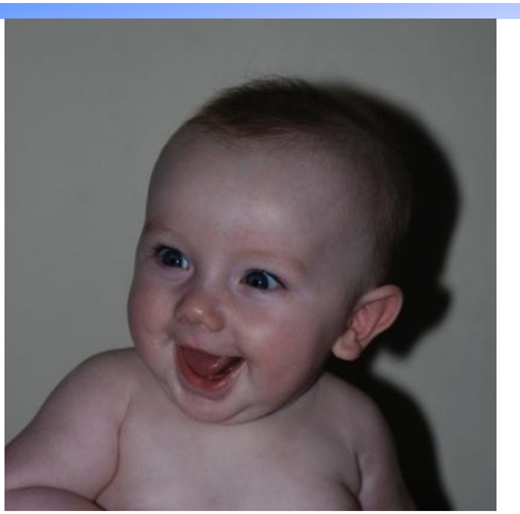
Minimum norm point algorithm: usually orders of magnitude faster

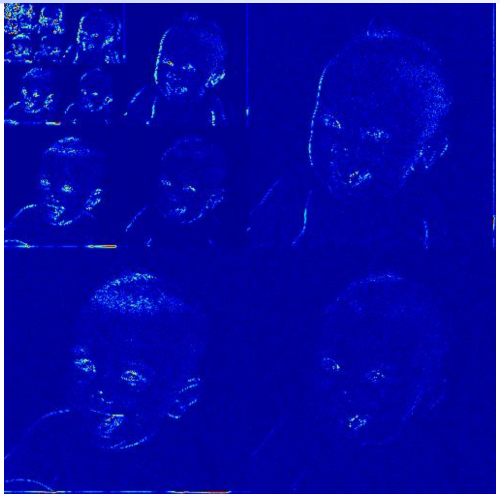
[Fujishige & Isotani ’11]

Example: Sparsity

d

pixels



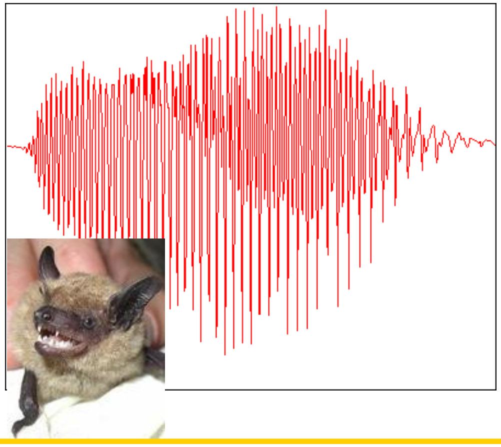


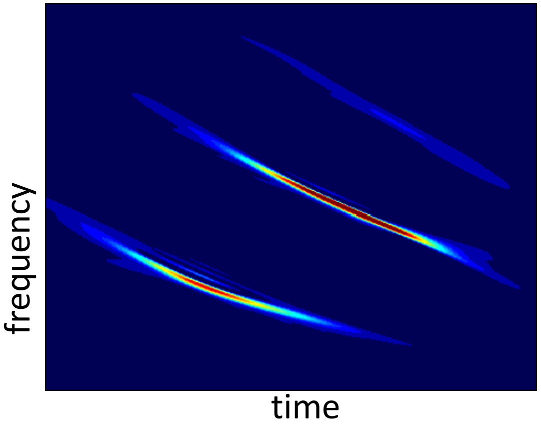
$k \ll d$

large wavelet coefficients

d

wideband signal samples



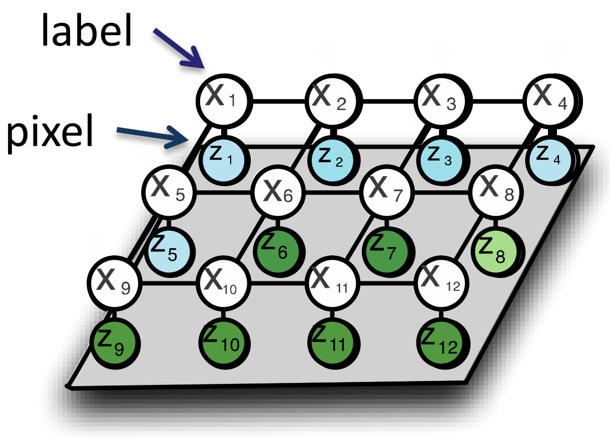


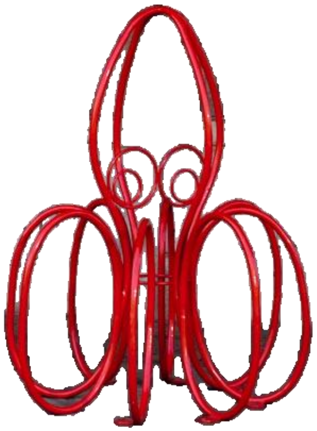
$k \ll d$

large Gabor (TF) coefficients

Many natural signals sparse in suitable basis.
Can exploit for learning/regularization/compressive sensing...

Example: MAP inference





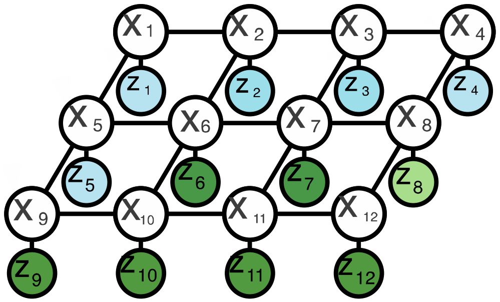
$$\max_{\mathbf{x}\in\{0,1\}^n} P(\mathbf{x} \mid \mathbf{z}) \propto \exp(-E(\mathbf{x};\mathbf{z}))$$

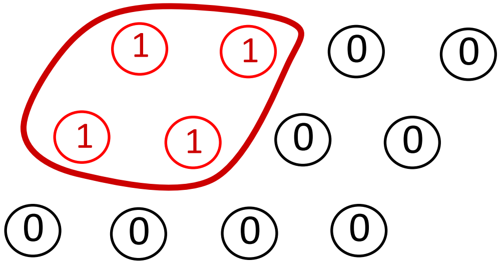
labels

pixel values

$$\Leftrightarrow \quad \min_{\mathbf{x}\in\{0,1\}^n} E(\mathbf{x};\mathbf{z})$$

Example: MAP inference





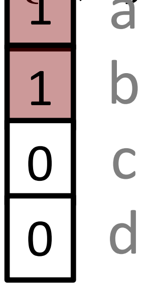
Recall: equivalence

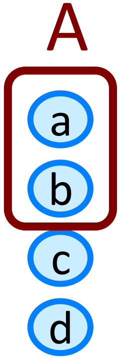
function on binary vectors $\qquad$ set function

$$\max_{\mathbf{x}\in\{0,1\}^n} P(\mathbf{x} \mid \mathbf{z}) \propto \exp(-E(\mathbf{x};\mathbf{z}))$$

$x = e_A$

$$E(e_A;\mathbf{z}) = F(A)$$





if F is submodular (attractive potentials), then

$$\Leftrightarrow \min_{\mathbf{x}\in\{0,1\}^n} E(\mathbf{x};\mathbf{z})$$

MAP inference = submodular minimization!

polynomial-time

Special cases

Minimizing general submodular functions:

poly-time, but not very scalable

Special structure → faster algorithms

- Symmetric functions
- Graph cuts
- Concave functions
- Sums of functions with bounded support
- ...

Fast approximate minimization

- Not all submodular functions can be optimized as graph cuts
- Even if they can: possibly many extra nodes in the graph ☹

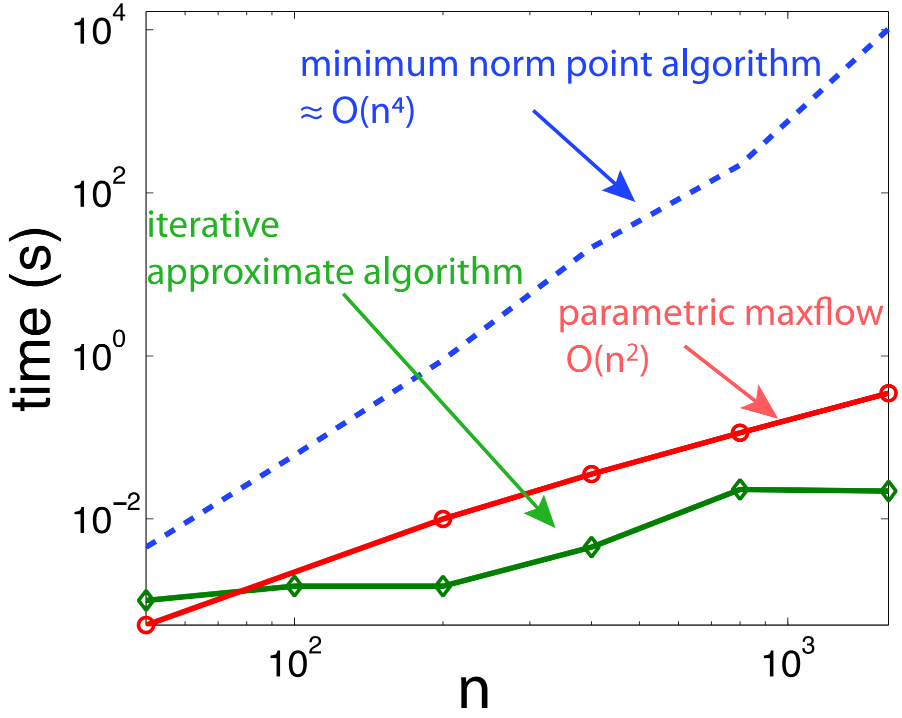
Other options?

- minimum norm algorithm
- other special cases:
  e.g. parametric maxflow

  [Fujishige & Iwata`99]

Approximate! ☺
Every submodular function can be approximated by a series of graph cut functions [Jegelka, Lin & Bilmes `11]

speech corpus selection [Lin&Bilmes `11]



Fast approximate minimization

- Not all submodular functions can be optimized as graph cuts
- Even if they can: possibly many extra nodes in the graph ☹

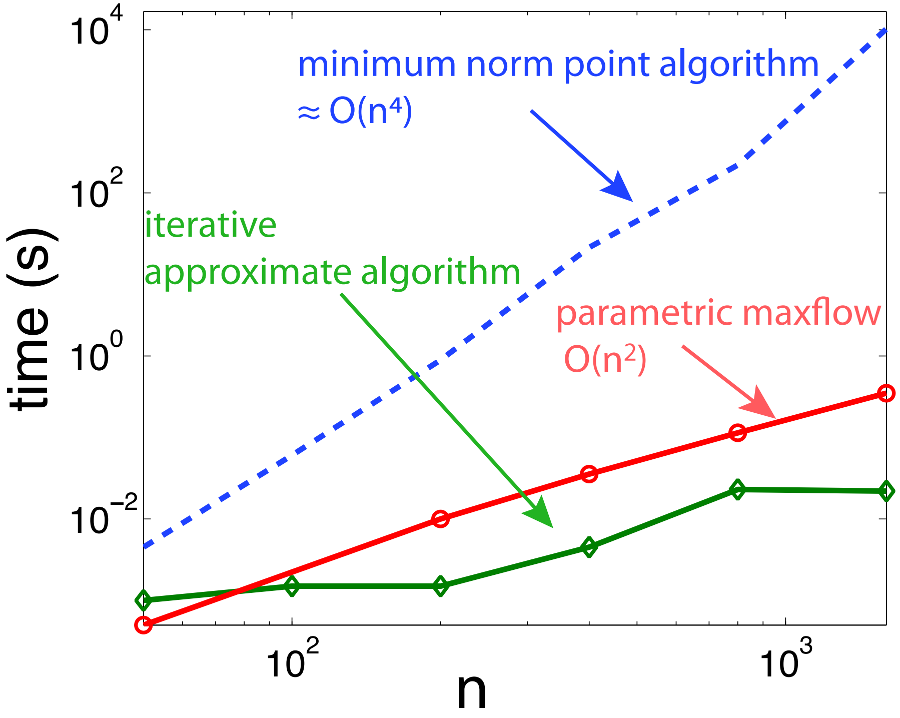
Approximate! ☺

decompose:

- represent as much as possible exactly by a graph
- rest: approximate iteratively by changing edge weights

solve a series of cut problems

speech corpus selection [Lin&Bilmes `11]



Other special cases

- Symmetric: $F(S) = F(V \setminus S)$
  - Queyranne‘s algorithm: $O(n^3)$

[Queyranne, 1998]

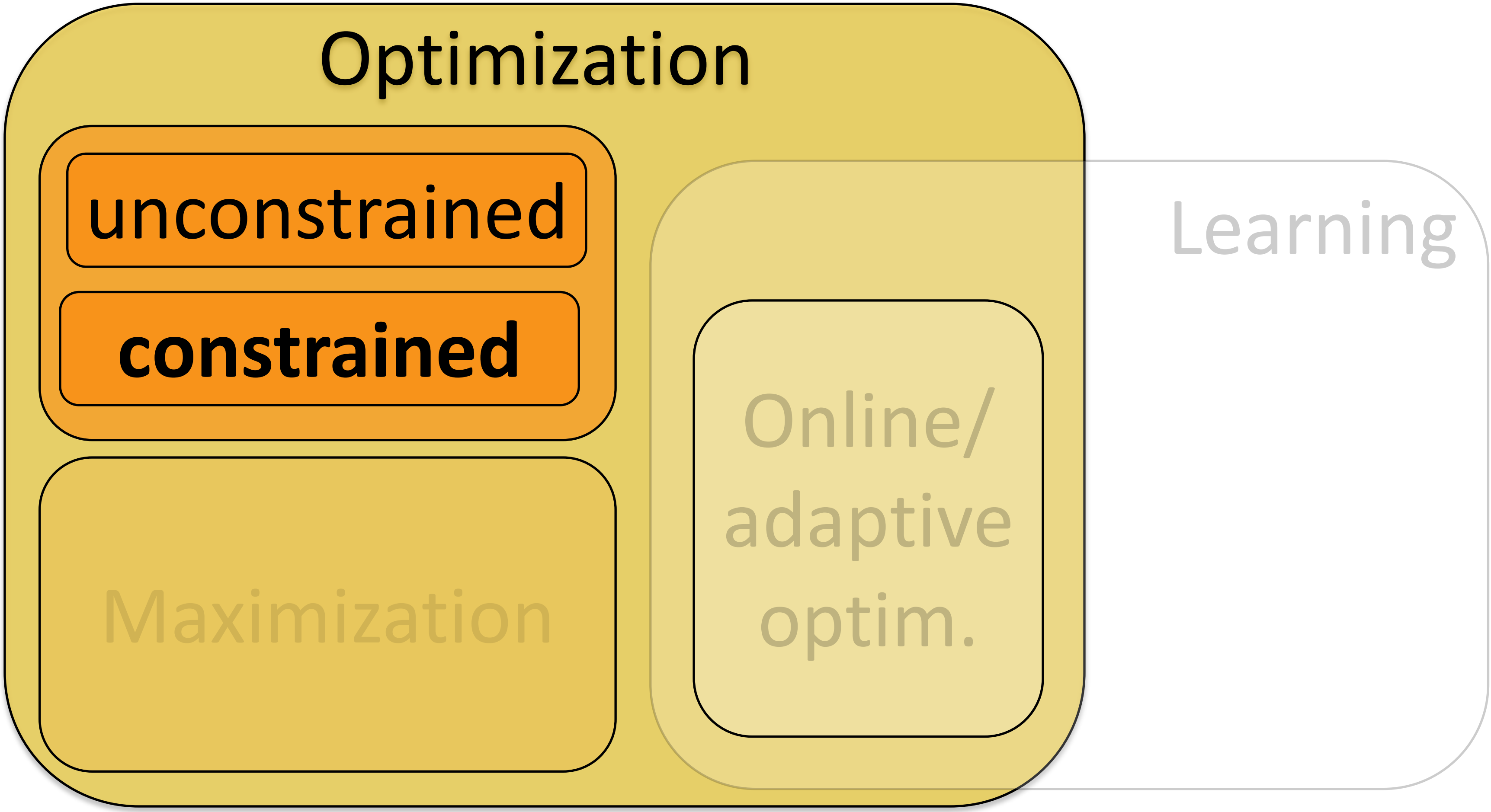
- Concave of modular: $F(S) = \sum_i g_i\Big(\sum_{s \in S} w(s)\Big)$

[Stobbe & Krause `10, Kohli et al, `09]

- Sum of submodular functions, each bounded support

[Kolmogorov `12]

Submodular minimization



Submodular minimization

- unconstrained: $\min \; F(A) \quad \text{s.t.} \; A \subseteq V$
  - nontrivial algorithms,
    polynomial time

special case:
balanced
cut

- constraints: e.g. $\min \; F(A) \quad \text{s.t.} \; |A| \geq k$
  - limited cases doable: $\dots$
    odd/even cardinality, inclusion/exclusion of a set

General case: NP hard

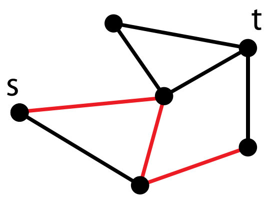
- hard to approximate within polynomial factors!
- But: special cases often still work well

[Lower bounds: Goel et al.`09, Iwata & Nagano `09, Jegelka & Bilmes `11]

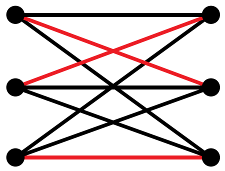
Constraints

minimum…

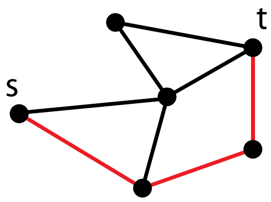
cut



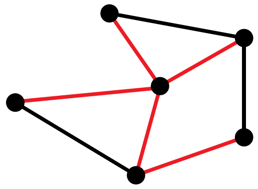
matching



path



spanning tree

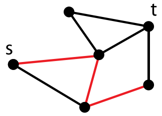


ground set: edges in a graph

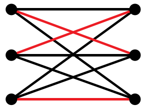
$$\min_{S\in\mathcal{C}} \sum_{e\in S} w(e) \quad \Longrightarrow \quad \min_{S\in\mathcal{C}} F(S)$$

Constrained optimization

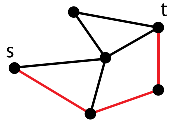
cut



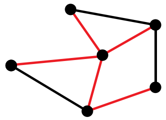
matching



path



spanning tree



$$\min_{S \in \mathcal{C}} \quad F(S)$$

approximate optimization

convex relaxation

minimize surrogate function

approximation bounds dependent on *F*:

polynomial – constant – FPTAS

$O(n)$ $(1+\epsilon)$

[Goel et al.`09, Iwata & Nagano `09, Goemans et al. `09, Jegelka & Bilmes `11, Iyer et al. ICML `13, Kohli et al `13...]

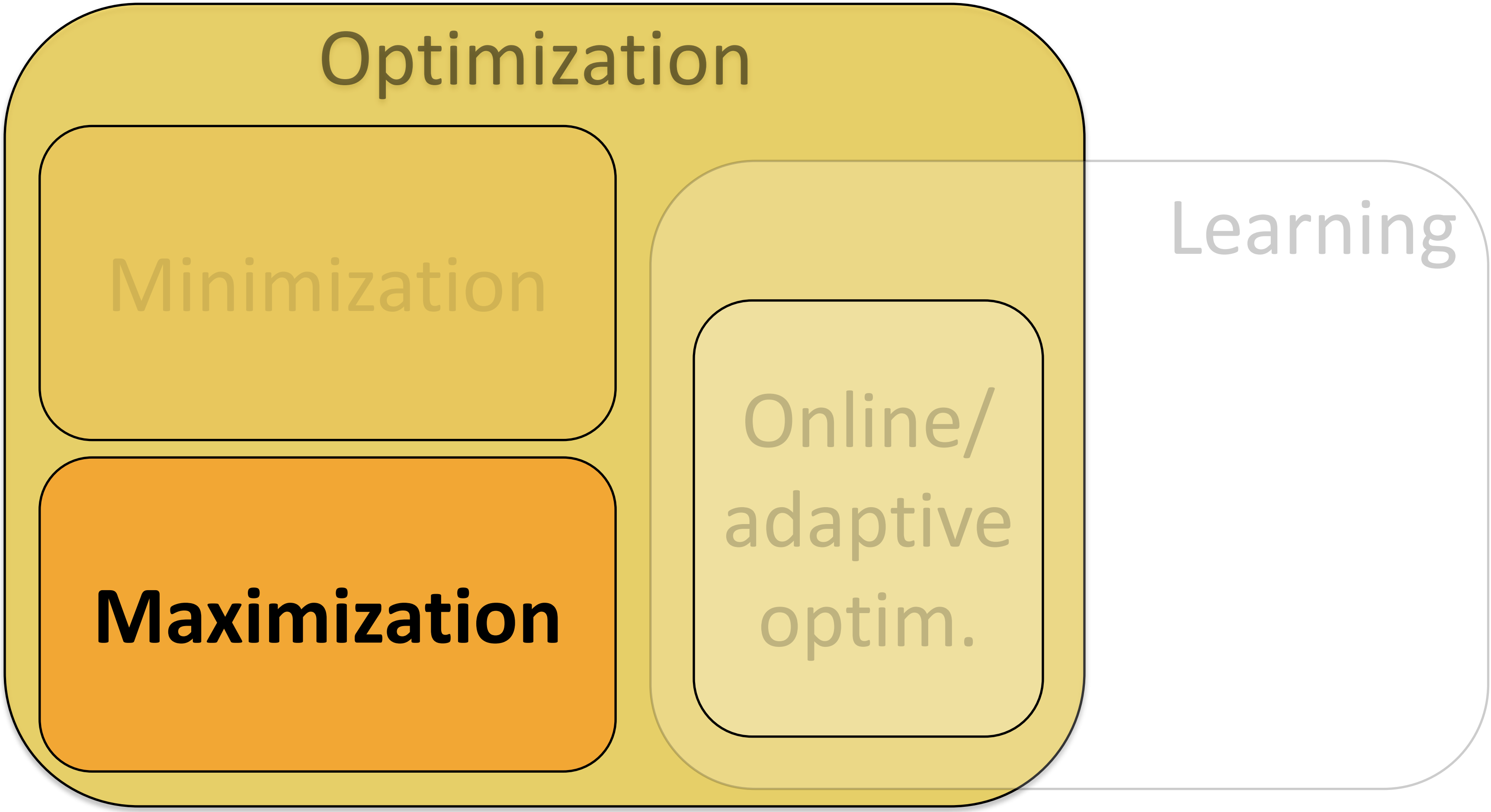
Submodular min in practice

- Does a special algorithm apply?
  - symmetric function? graph cut? .... approximately?
- Continuous methods: convexity
  - minimum norm point algorithm

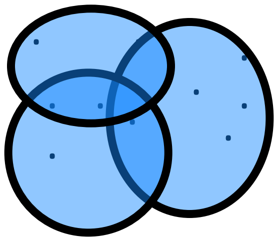
- Other techniques [not addressed here]
  - LP, column generation, …
- Combinatorial algorithms: relatively high complexity

- Constraints: hard
  - majorize-minimize or relaxation

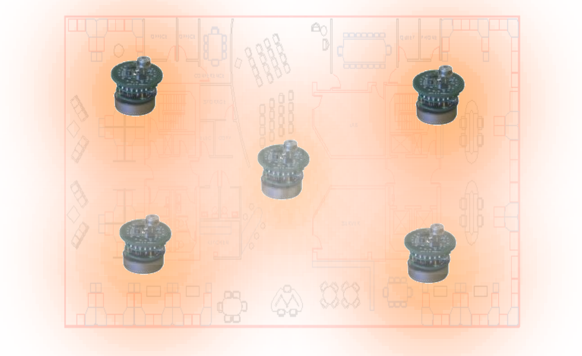
Optimization



Submodular maximization



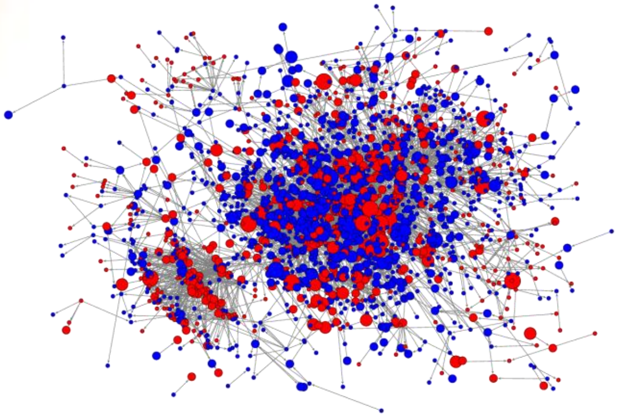
covering



sensing

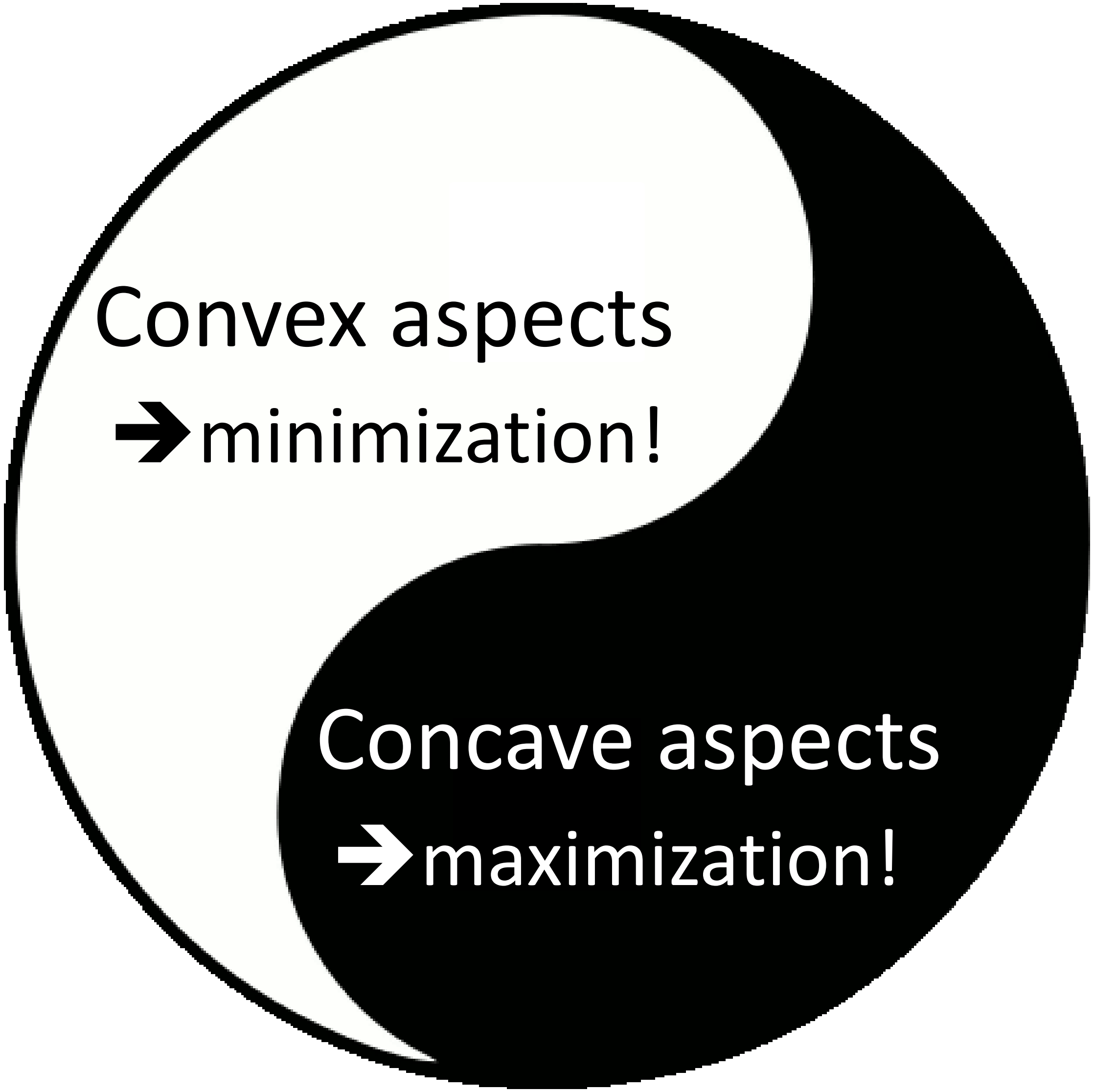
$$\max_{S \subseteq V} F(S)$$

summarization



network inference

Two faces of submodular functions



Convex aspects

→minimization!

Concave aspects

→maximization!

Submodular maximization

$$\max_{S \subseteq V} F(S)$$

→ submodularity and concavity

Concave aspects

- submodularity:

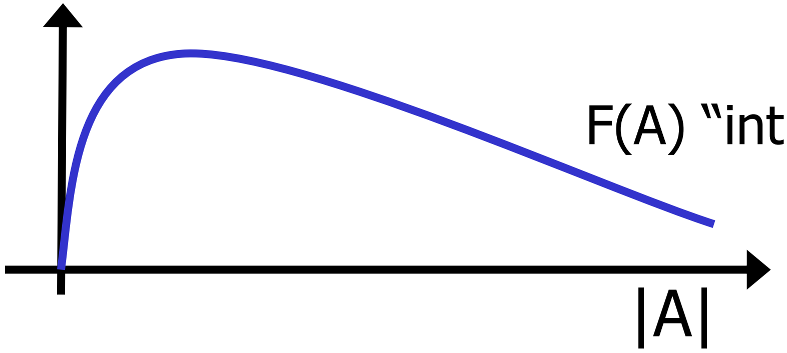
$$A \subseteq B, \quad s \notin B:$$

$$F(A \cup s) - F(A) \quad \geq \quad F(B \cup s) - F(B)$$

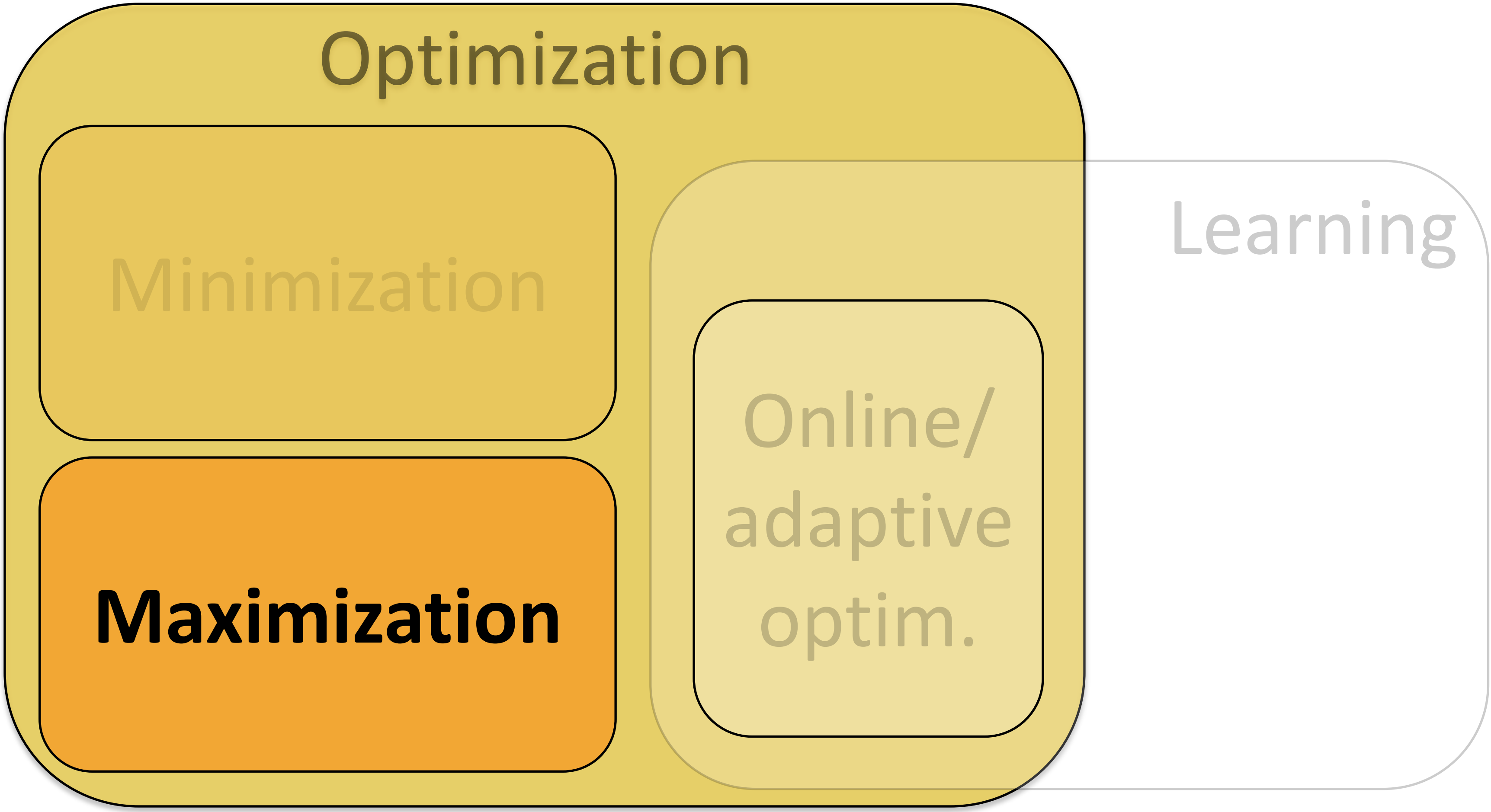
- concavity:

$$a \leq b, \quad s > 0:$$

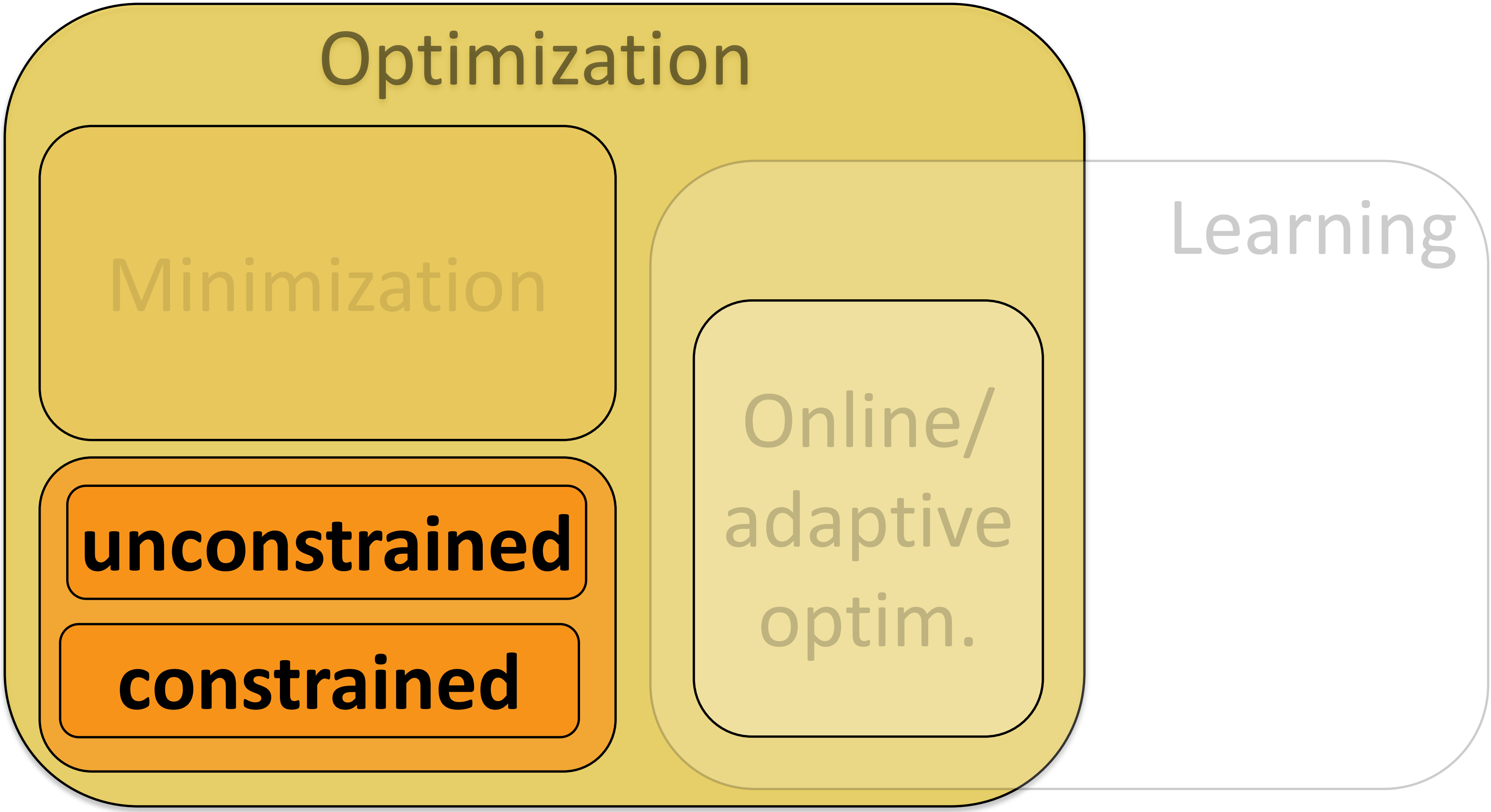
$$f(a+s) - f(a) \quad \geq \quad f(b+s) - f(b)$$



Optimization



Optimization



Optimization

Minimization

unconstrained

constrained

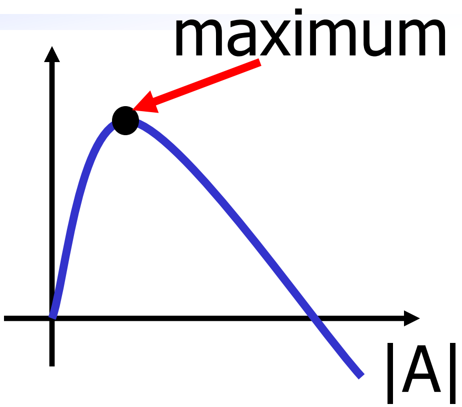
Learning

Online/
adaptive
optim.

Maximizing submodular functions

- Suppose we want for submodular F

$$A^* = \arg\max_A F(A) \text{ s.t. } A \subseteq V$$



- Example:
  - F(A) = U(A) – C(A) where U(A) is submodular utility, and C(A) is supermodular cost function

- In general: NP hard. Moreover:
- If F(A) can take negative values:
  As hard to approximate as maximum independent set (i.e., NP hard to get $O(n^{1-\varepsilon})$ approximation)

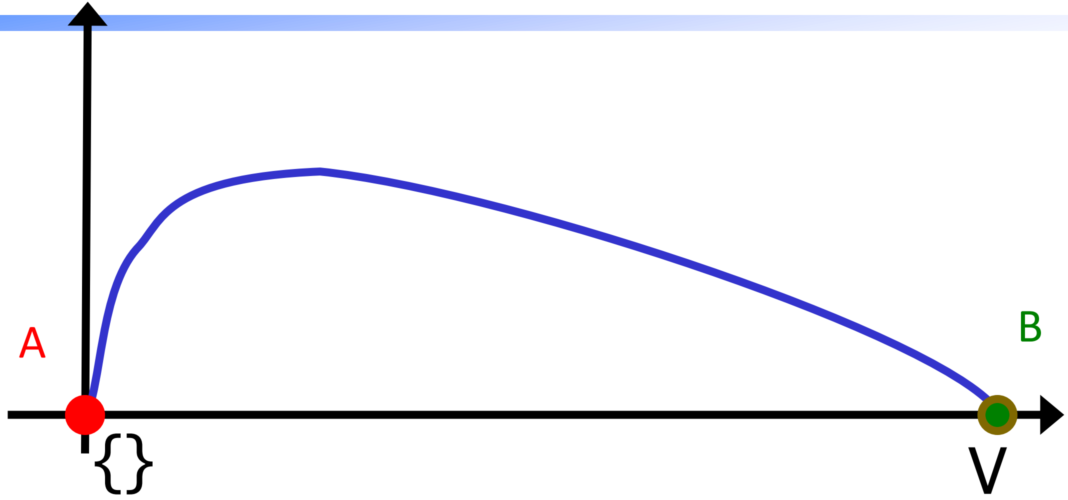
Exact maximization of SFs

- Mixed integer programming
  - Series of mixed integer programs [Nemhauser et al ‘81]
  - Constraint generation [Kawahara et al ‘09]
- Branch-and-bound
  - „Data-Correcting Algorithm“ [Goldengorin et al ’99]

Useful for small/moderate problems

All algorithms worst-case exponential!

Randomized USM (Buchbinder et al '12)



```
Start with A={}, B=V
For i=1 to n
    $v_+ = \max\Big(F(A \cup \{s_i\}) - F(A), 0\Big)$
    $v_- = \max\Big(F(B \setminus \{s_i\}) - F(B), 0\Big)$
    Pick $U \sim \mathrm{Unif}([0,1])$
    If $U \leq v_+/(v_+ + v_-)$ set $A \leftarrow A \cup \{s_i\}$
    Else $B \leftarrow B \setminus \{s_i\}$
Return $A \ (= B)$
```

Maximizing positive submodular functions

[Feige, Mirrokni, Vondrak ’09; Buchbinder, Feldman, Naor, Schwartz ’12]

Theorem

Given a nonnegative submodular function F,

RandomizedUSM returns set A_R such that

$$F(A_R) \geq 1/2 \ \max_A F(A)$$

- Cannot do better in general than ½ unless P = NP

Unconstrained vs. constraint maximization

Given monotone utility F(A) and cost C(A), optimize:

Option 1:

$$\boxed{\begin{aligned} &\max_A F(A) - C(A) \\ &\text{s.t. } A \subseteq V \end{aligned}}$$

"Scalarization"

Option 2:

$$\boxed{\begin{aligned} &\max_A F(A) \\ &\text{s.t. } C(A) \leq B \end{aligned}}$$

"Constrained maximization"

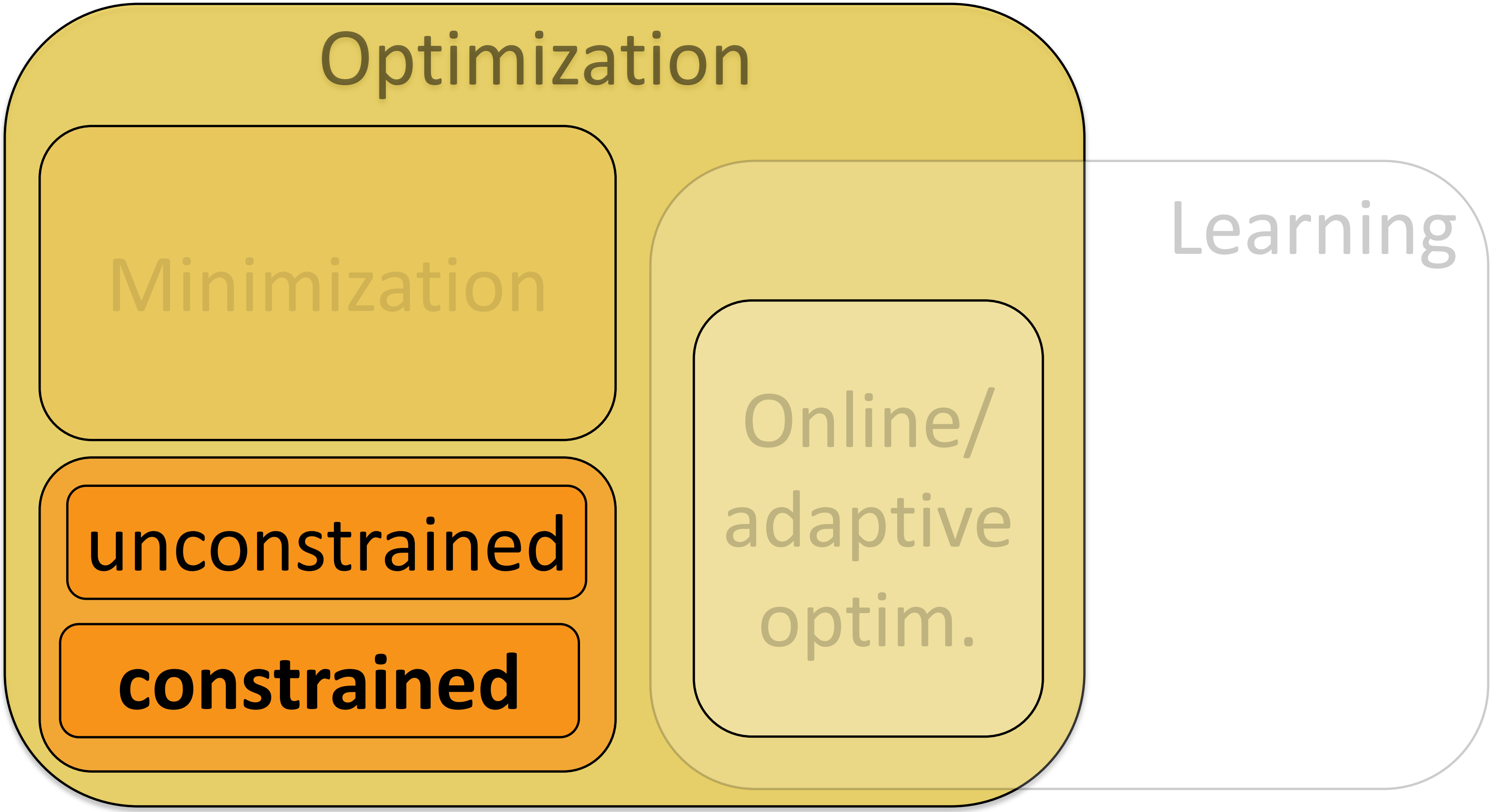
Can get 1/2 approx…

if F(A)-C(A) ≥ 0 for all sets A

Positiveness is a strong requirement ☹

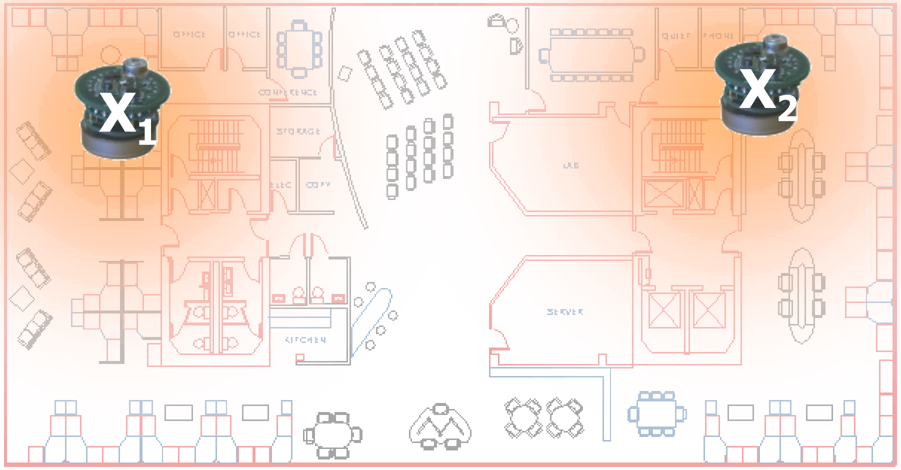
What is possible?

Optimization

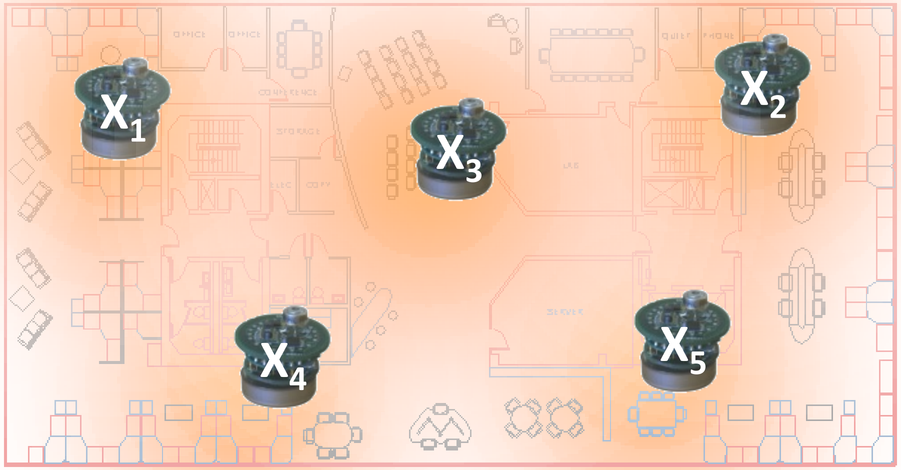


Monotonicity

Placement A = {1,2}



Placement B = {1,…,5}



F is monotonic:

$$\forall A, s : \underbrace{F(A \cup \{s\}) - F(A)}_{\Delta(s \mid A) \geq 0} \geq 0$$

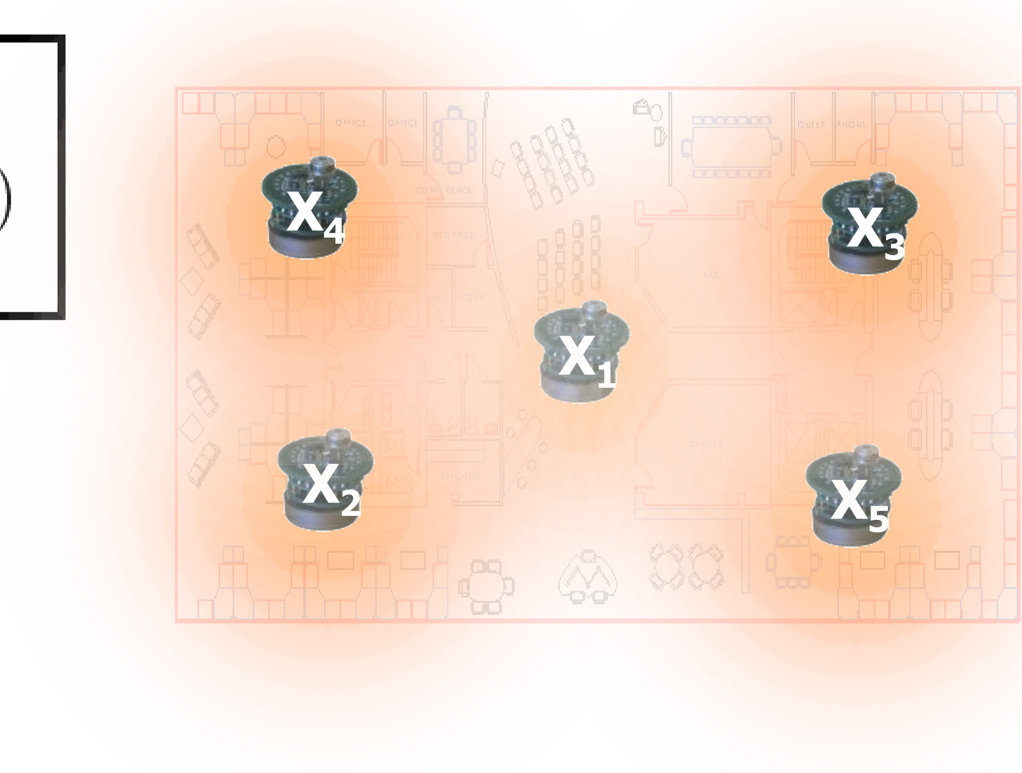
Adding sensors can only help

Cardinality constrained maximization

- **Given**: finite set V, monotone SF F
- **Want**:

$$\boxed{\mathcal{A}^* \subseteq \mathcal{V} \quad \text{such that} \\ \mathcal{A}^* = \operatorname*{argmax}_{|\mathcal{A}|\leq k} F(\mathcal{A})}$$

NP-hard!

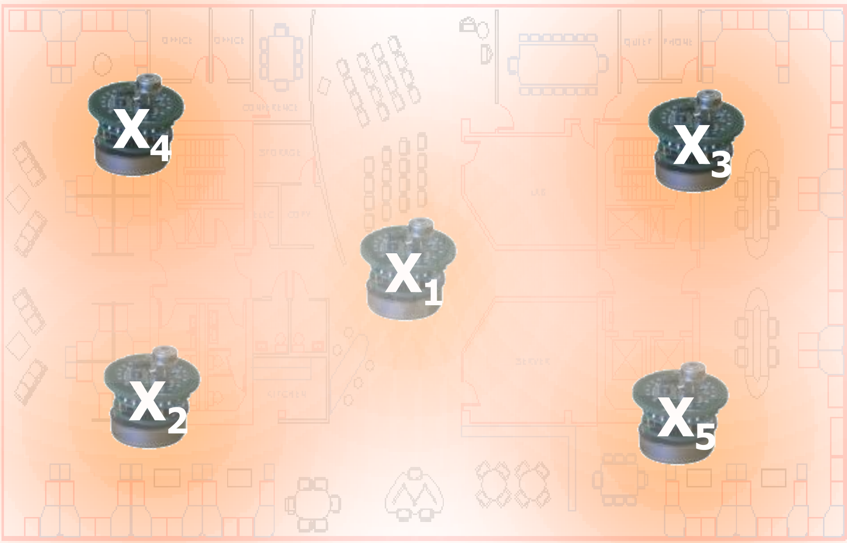


Greedy algorithm

- **Given**: finite set V, monotone SF F
- **Want**: $\mathcal{A}^* \subseteq \mathcal{V}$ such that

$$\mathcal{A}^* = \operatorname*{argmax}_{|\mathcal{A}| \leq k} F(\mathcal{A})$$

NP-hard!

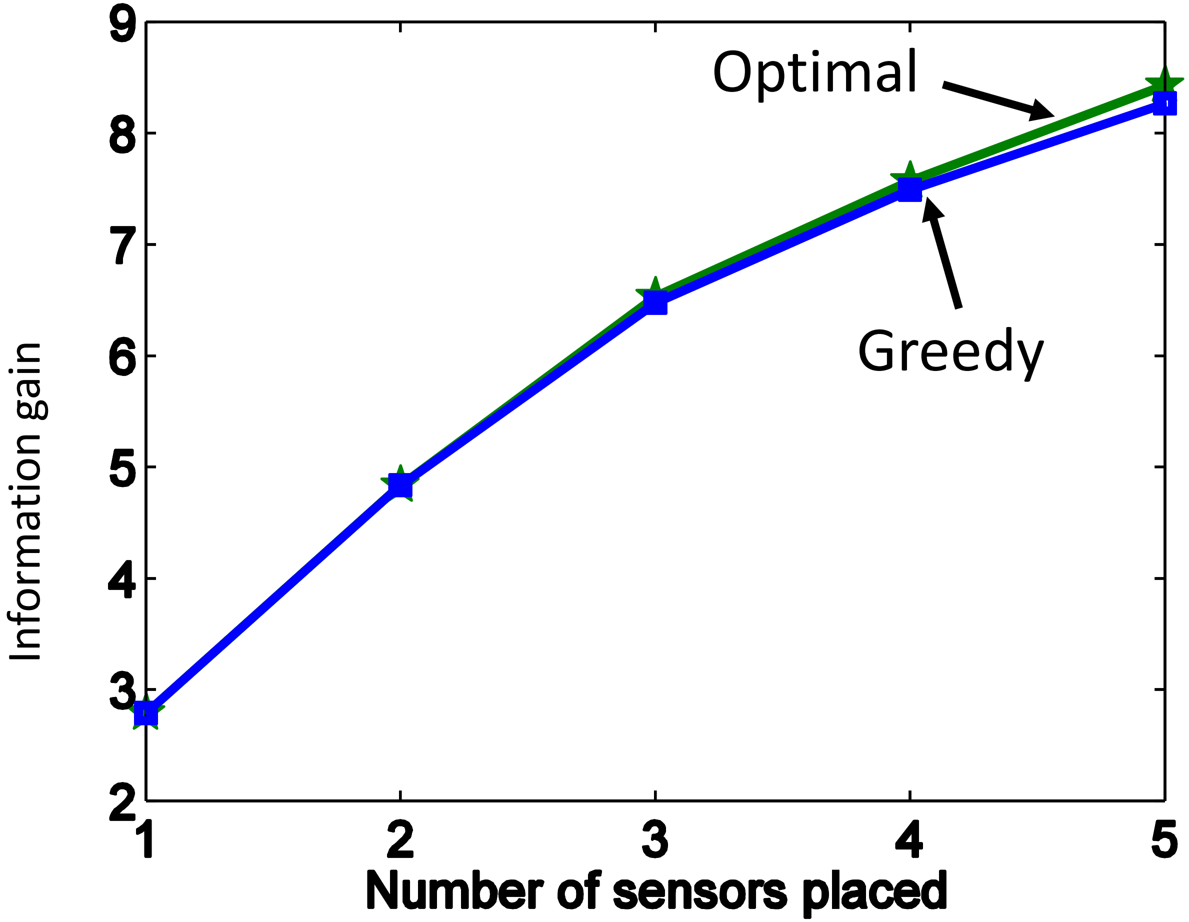


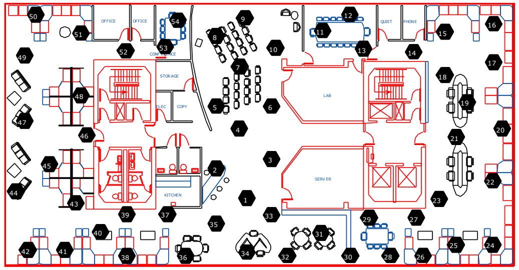
Greedy algorithm:

```
Start with $\mathcal{A} = \emptyset$
For i = 1 to k
    $s^* \leftarrow \arg\max_s F(\mathcal{A} \cup \{s\})$
    $\mathcal{A} \leftarrow \mathcal{A} \cup \{s^*\}$
```

How well can this simple heuristic do?

Performance of greedy





Temperature data
from sensor network

Greedy empirically close to optimal. Why?

One reason submodularity is useful

Theorem [Nemhauser, Fisher & Wolsey ’78]

For monotonic submodular functions,
Greedy algorithm gives constant factor approximation

$$F(A_{greedy}) \geq \underbrace{(1-1/e)}_{\sim 63\%} F(A_{opt})$$

- Greedy algorithm gives near-optimal solution!
- In general, need to evaluate exponentially many sets to do better! [Nemhauser & Wolsey ’78]
- Also many special cases are hard (set cover, mutual information, …)

Scaling up the greedy algorithm [Minoux ’78]

In round i+1,

- have picked $A_i = \{s_1,\ldots,s_i\}$
- pick $s_{i+1} = \mathrm{argmax}_s\ F(A_i\ \mathrm{U}\ \{s\})-F(A_i)$

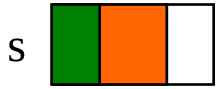
I.e., maximize “marginal benefit” $\otimes(s \mid A_i)$

$$\otimes(s \mid A_i) = F(A_i\ \mathrm{U}\ \{s\})-F(A_i)$$

Key observation: Submodularity implies

$$i \leq j \quad \Rightarrow \quad \otimes(s \mid A_i) \geq \otimes(s \mid A_j)$$

$\otimes(s \mid A_i) \geq \otimes(s \mid A_{i+1})$

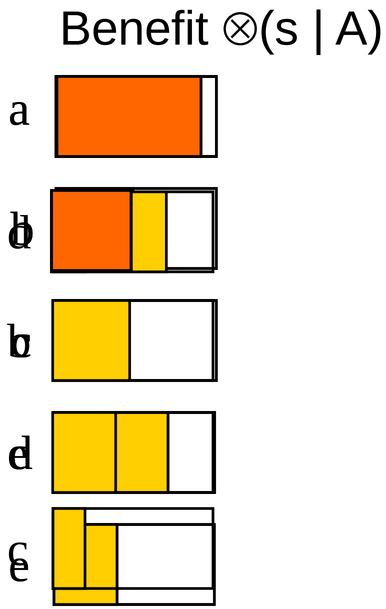


Marginal benefits can never increase!

“Lazy” greedy algorithm [Minoux ’78]

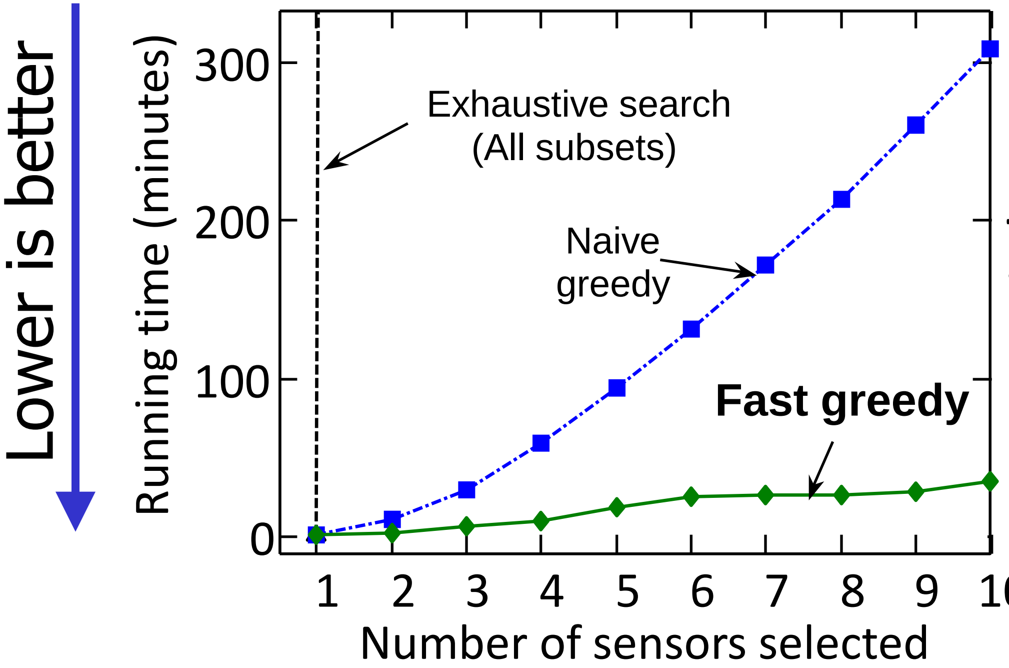
Lazy greedy algorithm:

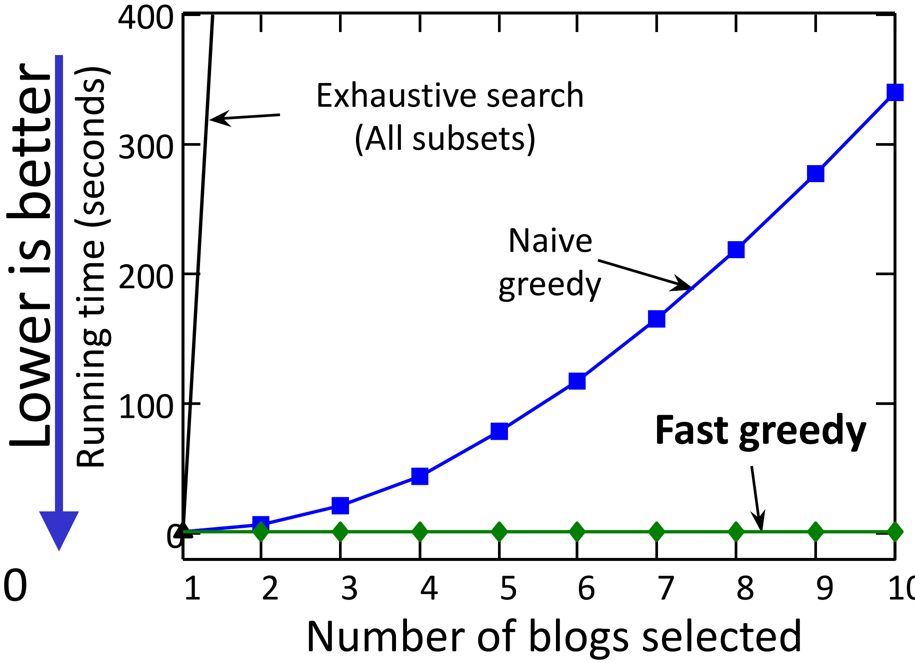
- First iteration as usual
- Keep an ordered list of marginal benefits $\otimes_i$ from previous iteration
- Re-evaluate $\otimes_i$ only for top element
- If $\otimes_i$ stays on top, use it, otherwise re-sort



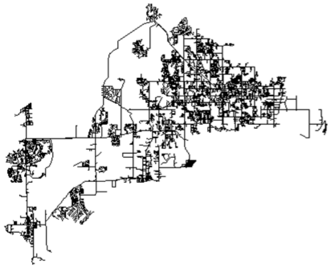
Note: Very easy to compute online bounds, lazy evaluations, etc. [Leskovec, Krause et al. ’07]

Empirical improvements [Leskovec, Krause et al'06]

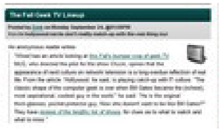




Sensor placement



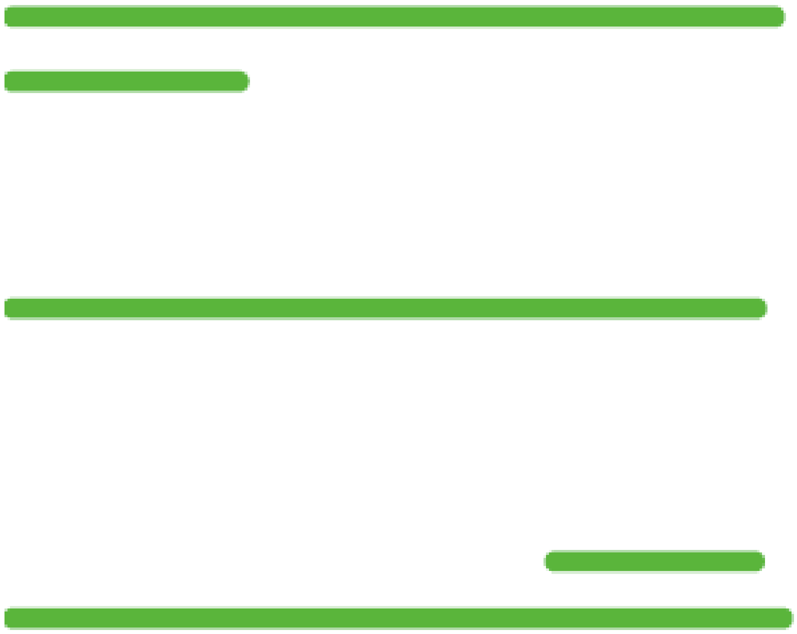
Blog selection



30x speedup

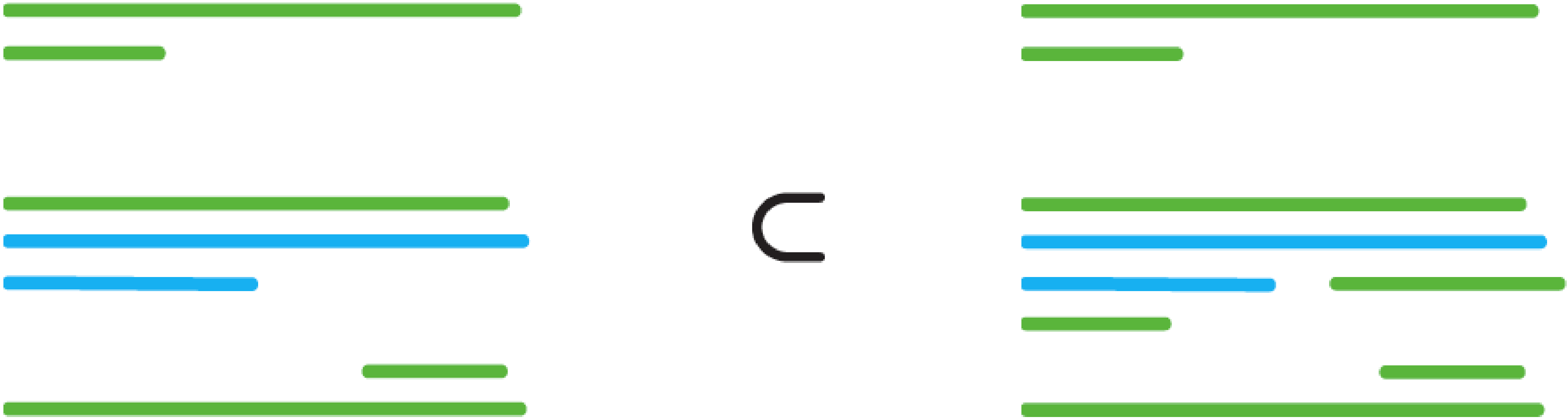
700x speedup

Document summarization [Lin & Bilmes ‘11]



- Which sentences should we select that best summarize a document?

Marginal gain of a sentence

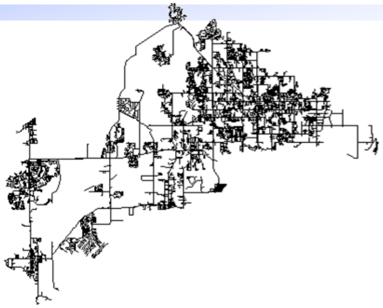


- Many natural notions of „document coverage“ are submodular [Lin & Bilmes ‘11]

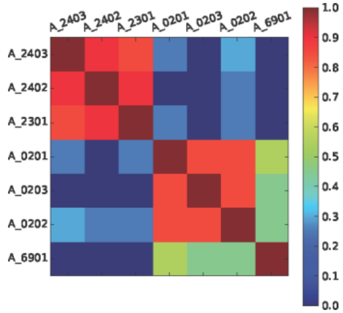
Submodular Sensing Problems

[with Guestrin, Leskovec, Singh, Sukhatme, …]

Environmental monitoring
[UAI’05, JAIR ’08, ICRA ‘10]



Water distribution networks
[J WRPM ’08]



Experiment design
[NIPS ‘10, ’11, PNAS’13]

Machine Learning (Theory)
Slashdot
boingboing A Directory Of Wonderful Things
sisu
engadget

Recommending blogs & news
[KDD ‘07, ’10]

Can all be reduced to monotonic submodular maximization

More complex constraints

- So far: $$\mathcal{A}^* = \operatorname*{argmax}_{|\mathcal{A}| \leq k} F(\mathcal{A})$$

- Can one handle more complex constraints?

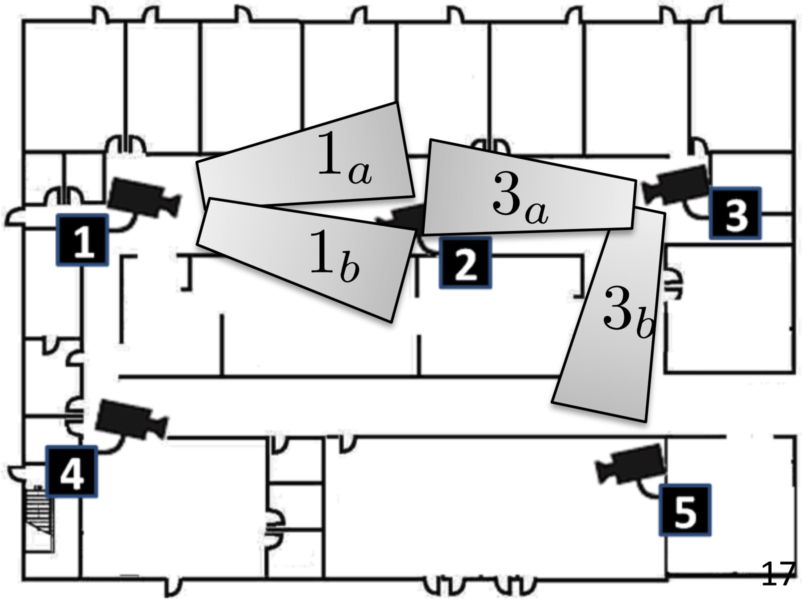
Example: Camera network

Ground set $V = \{1_a, 1_b, \dots, 5_a, 5_b\}$

Configuration: $S = \{v^1, \dots, v^k\}$

Sensing quality model $F : 2^V \to \mathbb{R}$

Configuration is feasible if no camera is pointed in two directions at once



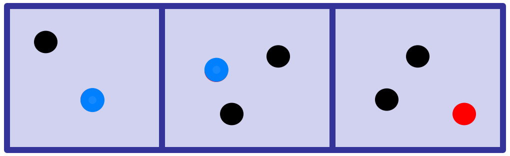
Matroids

- Abstract notion of feasibility: independence

S is independent if …

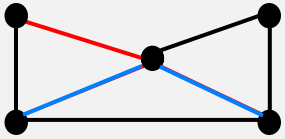
… |S| ≤ k

Uniform matroid



… S contains at most one element from each square

Partition matroid



… S contains no cycles

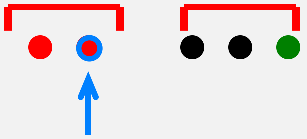
Graphic matroid

- S independent → $T \subseteq S$ also independent

Matroids

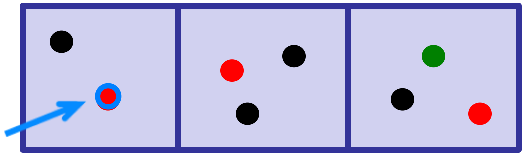
- Abstract notion of feasibility: independence

S is independent if …



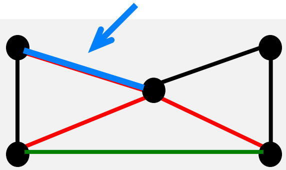
… |S| ≤ k

Uniform matroid



… S contains at most one element from each group

Partition matroid



… S contains no cycles

Graphic matroid

- S independent ➔ $T \subseteq S$ also independent
- Exchange property: S, U independent, $|S| > |U|$
  ➔ some $e \in S$ can be added to U: $U \cup e$ independent
- All maximal independent sets have the same size

Example: Camera network

Ground set $V = \{1_a, 1_b, \dots, 5_a, 5_b\}$

Configuration: $S = \{v^1, \dots, v^k\}$

Sensing quality model $F : 2^V \to \mathbb{R}$

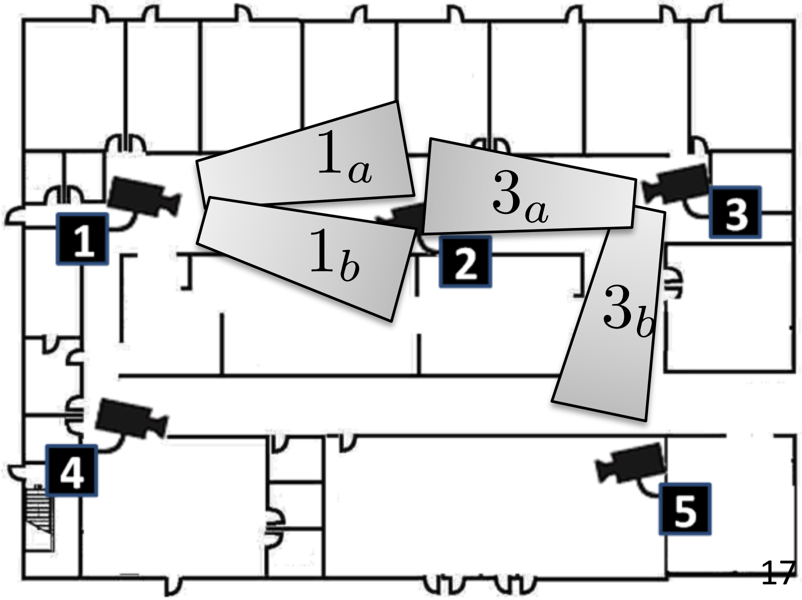
Configuration is feasible if no camera is pointed in two directions at once

This is a partition matroid:

$$P_1 = \{1_a, 1_b\}, \dots, P_5 = \{5_a, 5_b\}$$

Independence:

$$|S \cap P_i| \leq 1$$



Greedy algorithm for matroids:

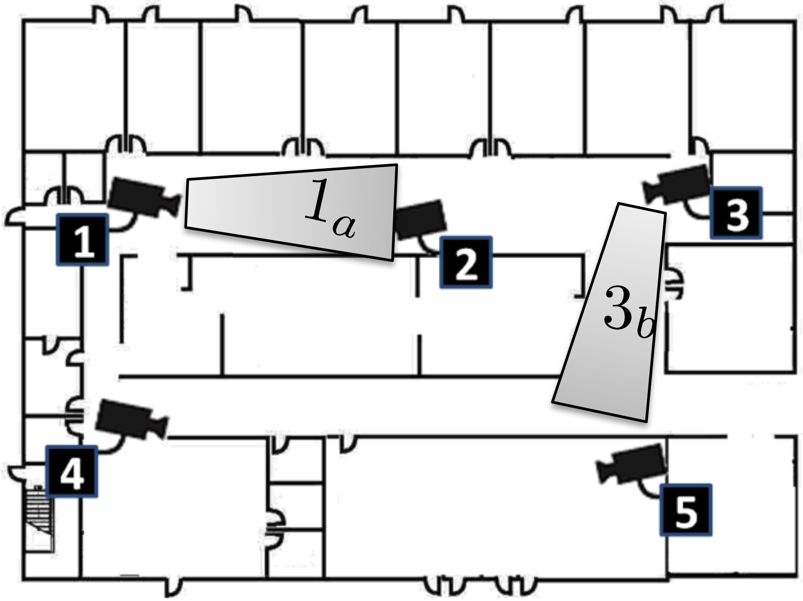
- Given: finite set V
- Want:

$\mathcal{A}^* \subseteq \mathcal{V}$ such that

$$\mathcal{A}^* = \operatorname*{argmax}_{A \text{ independent}} F(A)$$

Greedy algorithm:

```
    Start with $\mathcal{A} = \emptyset$
    While $\exists s : A \cup \{s\}$ indep.
        $s^* \leftarrow \operatorname*{argmax}_{s:\, A\cup\{s\} \text{ indep.}} F(A \cup \{s\})$
        $\mathcal{A} \leftarrow \mathcal{A} \cup \{s^*\}$
```



Maximization over matroids

Theorem [Nemhauser, Fisher & Wolsey ’78]

For monotonic submodular functions,
Greedy algorithm gives constant factor approximation

$$F(A_{greedy}) \geq \tfrac{1}{2}\ F(A_{opt})$$

- Greedy gives 1/(p+1) over intersection of p matroids
  - Can model matchings / rankings with p=2:
    Each item can be assigned ≤ 1 rank, each rank can take ≤ 1 item
- Can get also obtain (1-1/e) for arbitrary matroids [Vondrak et al ’08] using continuous greedy algorithm

Maximization: More complex constraints

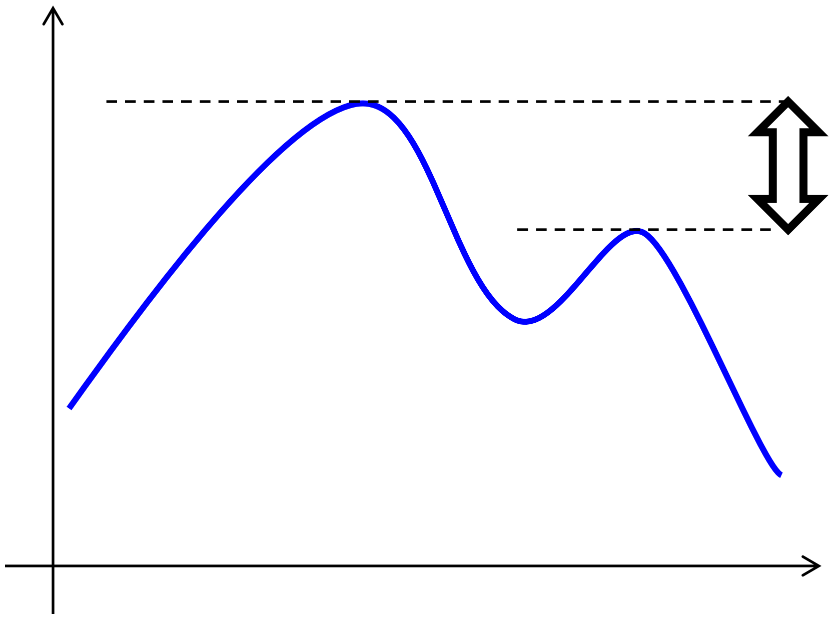
- Approximate submodular maximization possible under a variety of constraints:
  - (Multiple) matroid constraints
  - Knapsack (non-constant cost functions)
  - Multiple matroid and knapsack constraints
  - Path constraints (Submodular orienteering)
  - Connectedness (Submodular Steiner)
  - Robustness (minimax)
  - ...

Greedy works well

Need non-greedy algorithms

- Survey on „Submodular Function Maximization“ [Krause & Golovin ‘12] on submodularity.org

Key intuition for approx. maximization

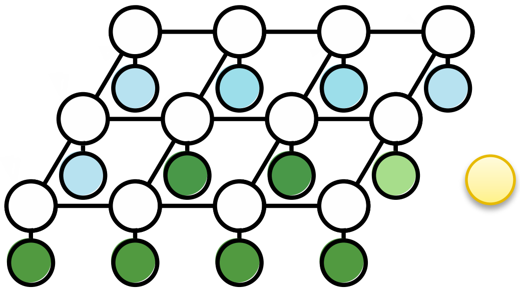


For submod. functions,
local maxima
can‘t be too bad

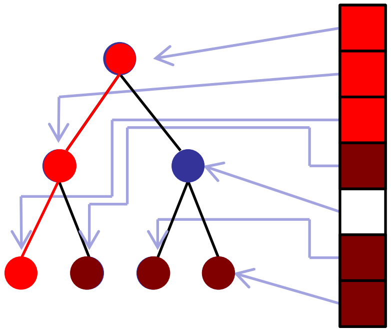
- E.g., all local maxima under cardinality constraints are within factor 2 of global maximum
- Key insight for more complex maximization
  ➔ Greedy, local search, simulated annealing for (non-monotone, constrained, ...)

Two-faces of submodular functions

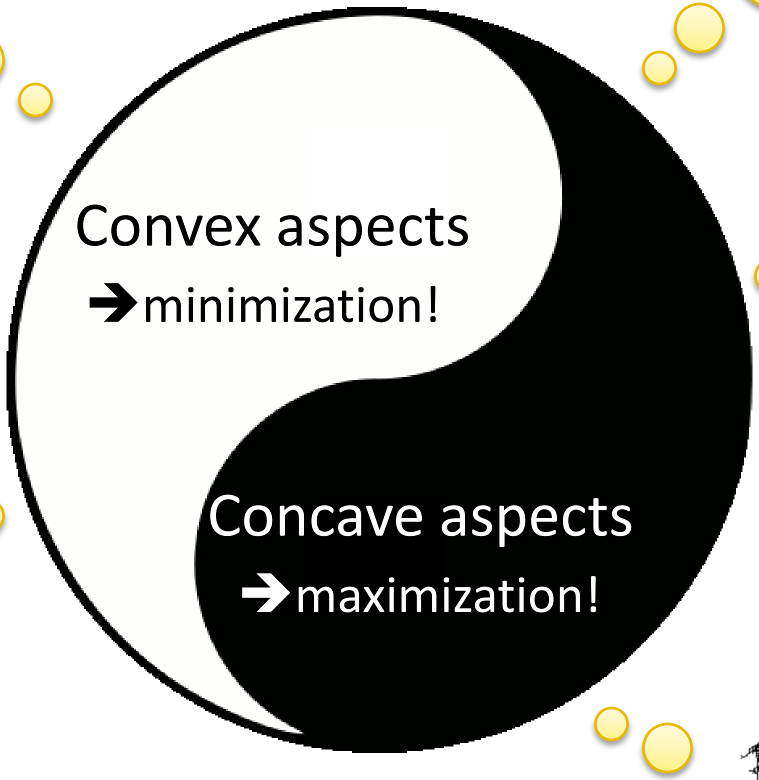
Cuts, clustering,
similarity



MAP inference



structured sparsity
regularization

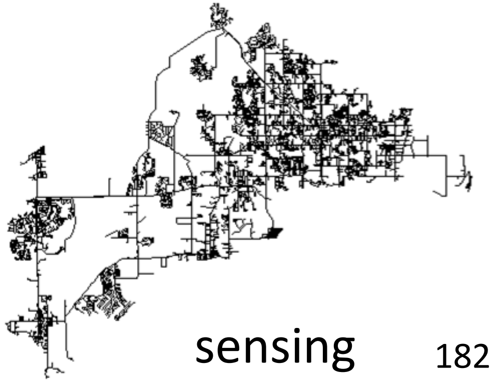


Convex aspects
➔minimization!

Concave aspects
➔maximization!

Coverage,
diversity

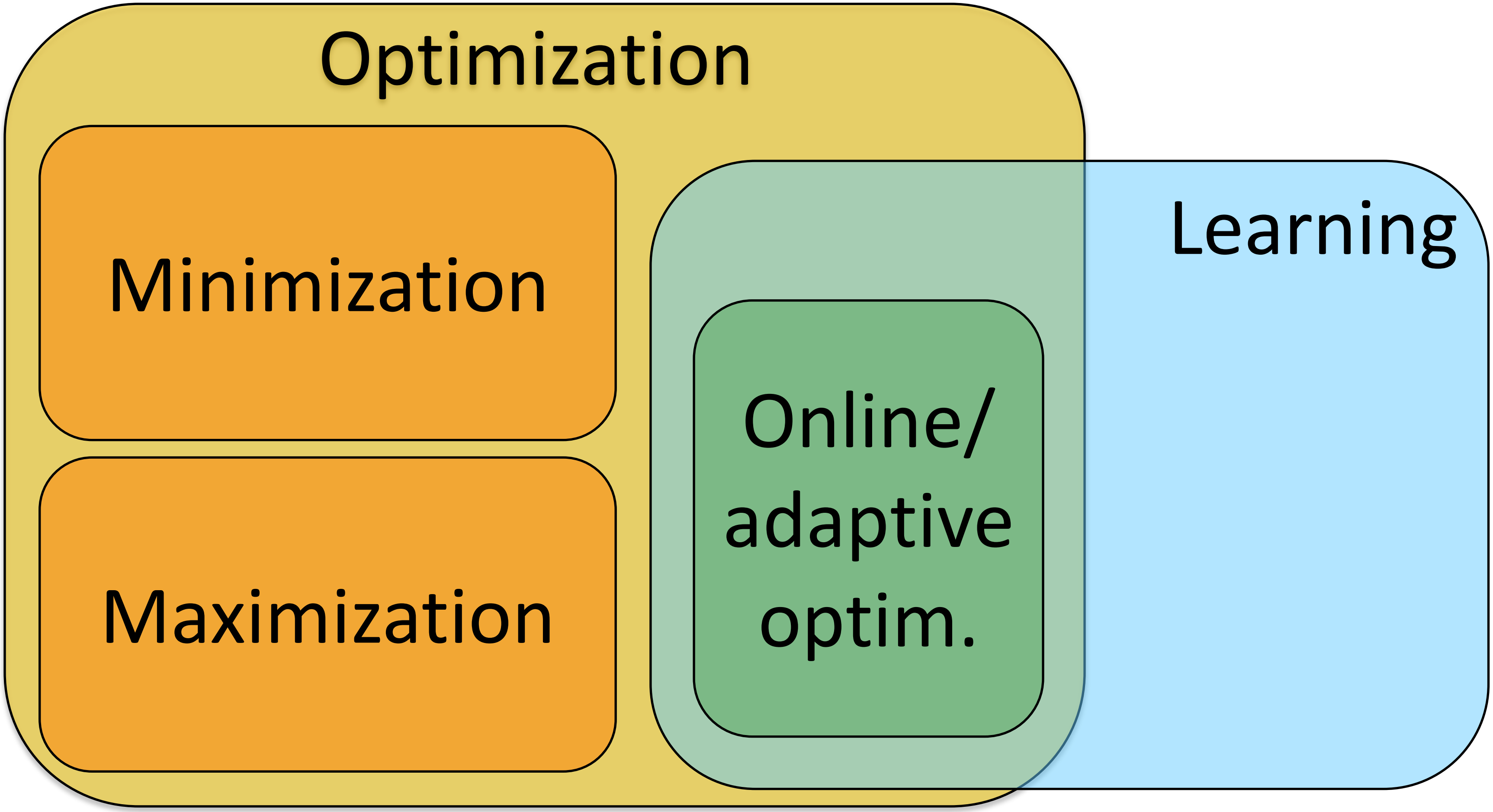
summarization



sensing

	Maximization	**Minimization**
Unconstrained	NP-hard, but well-approximable (if nonnegative)	Polynomial time! Generally inefficent (n^6), but can exploit special cases (cuts; symmetry; decomposable; ...)
Constrained	NP-hard but well-approximable „Greedy-(like)“ for cardinality, matroid constraints; Non-greedy for more complex (e.g., connectivity) constraints	NP-hard; hard to approximate, still useful algorithms

What to do with submodular functions



Optimization

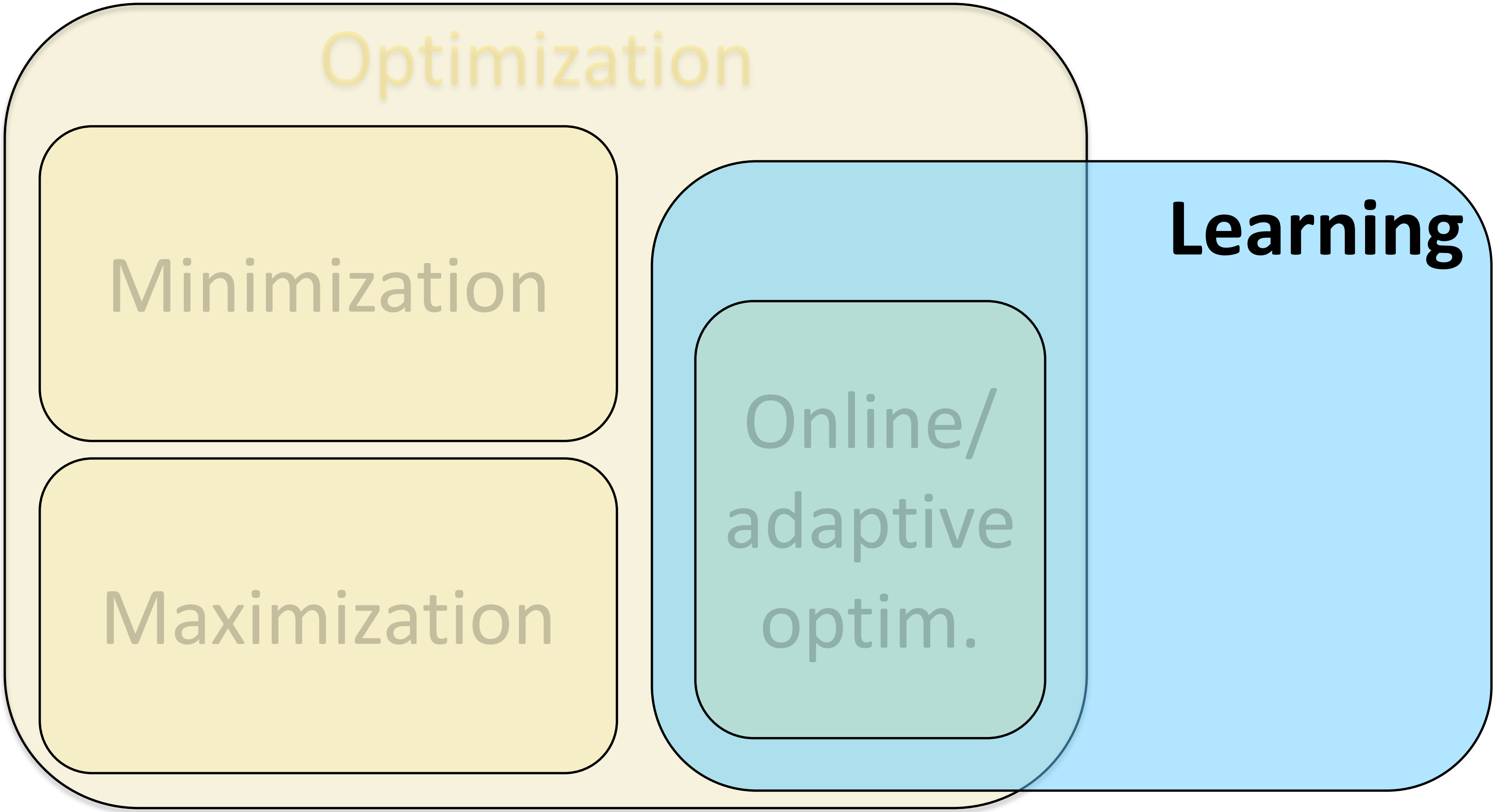
Minimization

Maximization

Learning

Online/ adaptive optim.

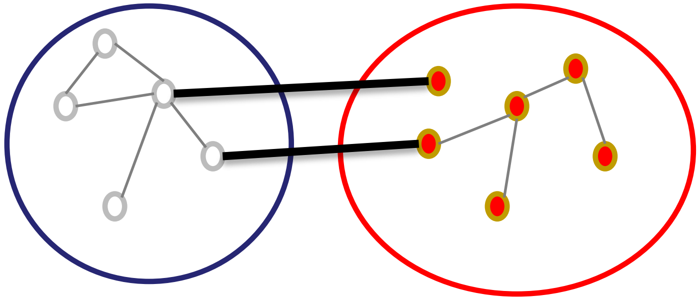
What to do with submodular functions



General Problem: Learning Set Functions

Base Set V

Set function $F : 2^V \to \mathbb{R}$



Can we learn F from few measurements / data?

$$\{(A_1, F(A_1)), \ldots, (A_m, F(A_m))\}$$

"Regressing" submodular functions

[Balcan, Harvey STOC '11]

- *Sample* m sets A_1 … A_m, from dist. D; see $F(A_1)$, …, $F(A_m)$
- From this, want to generalize well
- $\hat{F}$ is (α,ε,δ)-PMAC iff with prob. 1-δ it holds that

$$P_{A\sim\mathcal{D}}\left[\hat{F}(A) \leq F(A) \leq \alpha\hat{F}(A)\right] \geq 1-\varepsilon$$

Theorem: cannot approximate better than

$\alpha = n^{1/3} / \log(n)$

unless one looks at exponentially many samples A_i

But can efficiently obtain $\alpha = n^{1/2}$

Approximating submodular functions [Goemans, Harvey, Kleinberg, Mirrokni, ’08]

- *Pick* m sets, A_1 … A_m, get to see $F(A_1)$, …, $F(A_m)$
- From this, want to approximate F by $\hat{F}$ s.t.

$$\hat{F}(A) \leq F(A) \leq \alpha \hat{F}(A) \text{ for all A}$$

Theorem: Even if

- F is monotonic
- we can pick A_i adaptively,

cannot approximate better than $\alpha = n^{½} / \log(n)$ unless one looks at exponentially many sets A_i

But can efficiently obtain $\alpha = n^{½} \log(n)$

Other directions

- Game theory
  - Equilibria in cooperative (supermodular) games / fair allocations
  - Price of anarchy in non-cooperative games
  - Incentive compatible submodular optimization
- Generalizations of submodular functions
  - L#-convex / discrete convex analysis
  - XOS/Subadditive functions
- More optimization algorithms
  - Robust submodular maximization
  - Maximization and minimization under complex constraints
  - Submodular-supermodular procedure / semigradient methods
- Structured prediction with submodular functions

Further resources

- submodularity.org
  - Tutorial Slides
  - Annotated bibliography
  - Matlab Toolbox for Submodular Optimization
  - Links to workshops and related meetings
- discml.cc
  - NIPS Workshops on Discrete Optimization in Machine Learning
  - Videos of invited talks on videolectures.net

…

Conclusions

- Discrete optimization abundant in applications
- Fortunately, some of those have structure: submodularity
- Submodularity can be exploited to develop efficient, scalable algorithms with strong guarantees
- Can handle complex constraints
- Can learn to optimize (online, adaptive, …)

4th IEEE International Conference on Cognitive Infocommunications
CogInfoCom 2013
December 2-5
Budapest
HUNGARY

16:00 – 17:40 Tuesday Session 3 – Park II

Track: Cognitive capabilities of social networks

Session chair: Attila Kiss

16:00 Comparative study of Architecture for Twitter Analysis and a proposal for an improved approach - B. Molnár, Z. Vincellér

16:20 Towards Modeling Fuzzy Propagation for Sentiment Analysis in Online Social Networks: a Case study on TweetScope - D. N. Trung, J. J. Jung, L. A. Vu, A. Kiss

16:40 Five Ws, One H and Many Tweets - I. Szücs, G. Gombos, A. Kiss

17:00 On a Keyword-Lifecycle Model for Real-time Event Detection in Social Network Data - T. Matuszka, Z. Vincellér, S. Laki

17:20 Properties of the Most Influential Social Sensors - B. Kósa, B. Pinczel, G. Rácz, A. Kiss

Our publications in 2014

A basic network analytic package for RapidMiner

By Balázs Kósa, Márton Balassi, Péter Englert, Gábor Rácz, Zoltán Pusztai, Attila Kiss; Eötvös Lorand University

rapidminer

Location Schedule Buy Tickets

RapidMiner World
The All Data Analytics Conference
For All Enterprises

#RMWORLD • BOSTON, MA USA • AUGUST 18-21, 2014

Betweenness versus Linerank

By Balázs Kósa, Márton Balassi, Péter Englert, Attila Kiss

ICCCI 2014

Lecture Notes in Artificial Intelligence

Lecture Notes in Computer Science

6th International Conference on
Computational Collective Intelligence
Technologies and Applications
24th-26th September 2014, Seoul, Korea

An Improved Community-based Greedy Algorithm for Solving the Inuence Maximization Problem in Social Networks

By Gábor Rácz, Zoltán Pusztai, Balázs Kósa, Attila Kiss

ICAI 2014

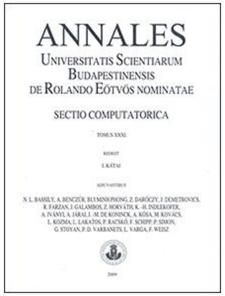
The 9th International Conference on Applied Informatics
to be held in Eger, Hungary
January 29–February 1, 2014

Quantitative analysis of Bitcoin exchange rate and transactional network properties

By Imre Szücs, Attila Kiss

Efficiency Issues of Computing Graph Properties of Social Networks

By Balázs Kósa, Márton Balassi, Péter Englert, Attila Kiss



ANNALES
Universitatis Scientiarum
Budapestinensis
de Rolando Eötvös nominatae

SECTIO COMPUTATORICA

TOMUS XXXI.

REDIGIT

I. KÁTAI

ADIUVANTIBUS

N. L. BASSILY, A. BENCZÚR, BUI MINH PHONG, Z. DARÓCZY, J. DEMETROVICS,
R. FARZAN, J. GALAMBOS, Z. HORVÁTH, K.-H. INDLEKOFER,
A. IVÁNYI, A. JÁRAI, J.-M. DE KONINCK, A. KÓSA, M. KOVÁCS,
L. KOZMA, L. LAKATOS, P. RACSKÓ, F. SCHIPP, P. SIMON,
G. STOYAN, P. D. VARBANETS, L. VARGA, F. WEISZ

2009

Community shells' effect on the disintegration dynamic of social networks

By Imre Szücs, Attila Kiss

Related Publications

- Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In **KDD’08**, pages 990-998, 2008.
- Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In **KDD’09**, pages 807-816, 2009.
- Chenhao Tan, Jie Tang, Jimeng Sun, Quan Lin, and Fengjiao Wang. Social action tracking via noise tolerant time-varying factor graphs. In **KDD’10**, pages 807–816, 2010.
- Lu Liu, Jie Tang, Jiawei Han, Meng Jiang, and Shiqiang Yang. Mining Topic-Level Influence in Heterogeneous Networks. In **CIKM’10**, pages 199-208, 2010.
- Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD’11**, pages 1397–1405, 2011.
- Jimeng Sun and Jie Tang. A Survey of Models and Algorithms for Social Influence Analysis. Social Network Data Analytics, Aggarwal, C. C. (Ed.), Kluwer Academic Publishers, pages 177–214, 2011.
- Jie Tang, Tiancheng Lou, and Jon Kleinberg. Inferring Social Ties across Heterogeneous Networks. In **WSDM'12**. pp. 743-752.
- Jia Jia, Sen Wu, Xiaohui Wang, Peiyun Hu, Lianhong Cai, and Jie Tang. Can We Understand van Gogh’s Mood? Learning to Infer Affects from Images in Social Networks. In **ACM MM**, pages 857-860, 2012.
- Lu Liu, Jie Tang, Jiawei Han, and Shiqiang Yang. Learning Influence from Heterogeneous Social Networks. In **DMKD**, 2012, Volume 25, Issue 3, pages 511-544.
- Jing Zhang, Biao Liu, Jie Tang, Ting Chen, and Juanzi Li. Social Influence Locality for Modeling Retweeting Behaviors. In **IJCAI'13**.
- Jie Tang, Sen Wu, and Jimeng Sun. Confluence: Conformity Influence in Large Social Networks. In **KDD'2013**.
- Jimeng Sun and Jie Tang. Models and Algorithms for Social Influence Analysis. In **WSDM’13**. (Tutorial)
- Tiancheng Lou, Jie Tang, John Hopcroft, Zhanpeng Fang, Xiaowen Ding. Learning to Predict Reciprocity and Triadic Closure in Social Networks. In **TKDD**, 2013.

References

- N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In **WSDM’08**, pages 207–217, 2008.
- A. Anagnostopoulos, R. Kumar, M. Mahdian. Influence and correlation in social networks. In **KDD’08**, pages 7-15, 2008.
- S. Aral, L. Muchnik, and A. Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. **PNAS**, 106 (51):21544-21549, 2009.
- S. Aral and D Walker. Identifying Influential and Susceptible Members of Social Networks. **Science**, 337:337-341, 2012.
- Barabasi and Albert (1999). Emergence of scaling n complex networks.
- E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In **EC ’09**, pages 325–334, New York, NY, USA, 2009. ACM.
- E. Bakshy, D. Eckles, R. Yan, and I. Rosenn. Social influence in social advertising: evidence from field experiments. In **EC'12**, pages 146-161, 2012.
- P. Bonacich. Power and centrality: a family of measures. **American Journal of Sociology**, 92:1170–1182, 1987.
- R. M. Bond, C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle and J. H. Fowler. A 61-million-person experiment in social influence and political mobilization. **Nature**, 489:295-298, 2012.
- R. S. Burt. Structural holes and good ideas. American Journal of Sociology, 110:349–399, 2004.
- W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In **KDD'09**, pages 199-207, 2009.

References(cont.)

- R. B. Cialdini and N. J. Goldstein. Social influence: compliance and conformity. **Annu Rev Psychol**, 55:591–621, 2004.
- D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In **KDD'08**, pages 160–168, 2008.
- P. Domingos and M. Richardson. Mining the network value of customers. In **KDD'01**, pages 57–66, 2001.
- R. Dunbar. Neocortex size as a constraint on group size in primates. **Human Evolution**, 1992, 20: 469–493.
- P. W. Eastwick and W. L. Gardner. Is it a game? evidence for social influence in the virtual world. **Social Influence**, 4(1):18–32, 2009.
- S. M. Elias and A. R. Pratkanis. Teaching social influence: Demonstrations and exercises from the discipline of social psychology. **Social Influence**, 1(2):147–162, 2006.
- Erdős, P.; Rényi, A. (1959), "On Random Graphs.".
- T. L. Fond and J. Neville. Randomization tests for distinguishing social influence and homophily effects. In **WWW'10**, 2010.
- J.H. Fowler and N.A. Christakis. The Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study. **British Medical Journal** 2008; 337: a2338
- M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring Networks of Diffusion and Influence. In **KDD'10**, pages 1019–1028, 2010.
- A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In **CIKM'08**, pages 499–508, 2008.
- A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In **WSDM'10**, pages 207–217, 2010.

References(cont.)

- G. Jeh and J. Widom. Scaling personalized web search. In **WWW '03**, pages 271-279, 2003.
- G. Jeh and J. Widom, SimRank: a measure of structural-context similarity. In **KDD'02**, pages 538-543, 2002.
- D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In **KDD'03**, pages 137–146, 2003.
- J. Kleinberg. Authoritative sources in a hyperlinked environment. **Journal of the ACM**, 46(5):604–632, 1999.
- Lazarsfeld et al. (1944). The people's choice: How the voter makes up his mind in a presidential campaign.
- J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In **KDD'07**, pages 420–429, 2007.
- S. Milgram. The Small World Problem. **Psychology Today**, 1967, Vol. 2, 60–67
- R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. Science, 2004
- http://klout.com
- P. Moore. Why I Deleted My Klout Profile, at **Social Media Today**, originally published November 19, 2011; retrieved November 26 2011
- M. E. J. Newman. A measure of betweenness centrality based on random walks. **Social Networks**, 2005.
- L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report SIDL-WP-1999-0120, Stanford University, 1999.
- D. B. Rubin, 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of Educational Psychology** 66, 5, 688–701.

References(cont.)

- J. Scripps, P.-N. Tan, and A.-H. Esfahanian. Measuring the effects of preprocessing decisions and network forces in dynamic network analysis. In **KDD'09**, pages 747–756, 2009.
- J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. In **ICDM'05**, pages 418–425, 2005.
- J. Ugandera, L. Backstromb, C. Marlowb, and J. Kleinberg. Structural diversity in social contagion. **PNAS**, 109 (20):7591-7592, 2012.
- D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. **Nature**,393(6684), pages 440–442, Jun 1998.
- http://en.wikipedia.org/wiki/Randomized_experiment

Thank you...
Thank You
Thanks
Thank you!