

Submodular Function Optimization

<http://bicmr.pku.edu.cn/~wenzw/bigdata2018.html>

Acknowledgement: this slides is based on Prof. Andreas Krause's, Prof. Jeff Bilmes, Prof. Francis Bach and Prof. Shaddin Dughmi lecture notes

Outline

- 1 What is submodularity?
 - Examples in recommendation sets
 - Definition
- 2 Submodular maximization
- 3 Submodular minimization
- 4 Applications of submodular maximization

Outline

- 1 What is submodularity?
 - Examples in recommendation sets
 - Definition
- 2 Submodular maximization
- 3 Submodular minimization
- 4 Applications of submodular maximization

Case study: News article recommendation

Home Mail Search News Sports Finance Weather Games Answers Screen Flickr Mobile More ▾

YAHOO!
NEWS

Mother's Day Gifts Lucky Brand handbags Ecco sandals All-inclusive vacations

Search

Sign In Mail More

News Home

U.S.
World
Politics
Tech
Science
Health
Odd News
Local
Dear Abby
Comics
ABC News
Yahoo Originals
Photos

State of emergency declared as Baltimore riots
Rioters looted stores and hurled rocks and bricks at police.
Read More ▾
▶ Raw: Protestors and Police in Baltimore
🕒 Protests turn violent in wake of Freddie Gray death

Parents of Colorado theater shooting victim fear copycat massacre
Loretta Lynch sworn in as new U.S. attorney general

All News Yahoo Originals News AP Reuters

US to speed up warhead dismantling, issues warning to N. Korea
The United States on Monday announced plans to step up the dismantling of retired nuclear warheads and issued a stern warning to North Korea to give up its nuclear program or face deeper isolation. US Secretary of State John Kerry also said world powers AFP

Second Navy SEAL dies after Virginia pool training accident
(Reuters) - A second U.S. Navy SEAL has died after a mishap while training in a swimming pool at a Virginia military base, according to media reports on Monday. Both members of the Navy's elite sea, Reuters

Woman convicted of murder after young daughter disappeared
PHOENIX (AP) — An Arizona mother was convicted of murder Monday in the death of her 5-year-old daughter, who prosecutors say was beaten, neglected and confined to a closet before being dumped in Associated Press

Detectives: Cumberland County woman raped, robbed, and run over
Sheriff's detectives said a woman was raped, robbed, run over, by men she knew from a school for the deaf

YAHOO! NEWS Must Watch

Raw: Looters Target Stores in Baltimore
Associated Press Video

Katie Couric
Global News Anchor
The Armenian genocide explained
Remembering the Oklahoma City bombing 20 years later

Matt Bai
National Political Columnist
Lincoln Chafee's 2016 run might matter more than...
Chris Christie bets on bold
Two friends. Two principles. One answer on Iran.

Michael Isikoff
Chief Investigative Correspondent
Facing questions over foreign donations, Hillary...
The mystery moneymen behind Ted Cruz's su

Case study: News article recommendation

Baijiahao News

新闻 网页 贴吧 知道 音乐 图片 视频 地图 百科 文库

百度一下 帮助 | 高级搜索 | 设置

新闻全文 新闻标题

首页 百家 个性推荐 互联网 娱乐 军事 财经 体育 国内 国际 社会 科技 女人 汽车 房产 游戏 教育 视频 名站

华西都市报 澎湃 南方都市报 参考消息 南方周末 财新网 北京青年报 21世纪经济报道 更多

习近平首次公开发表全面依法治国部分论述

李克强会见微宗顾先生 张德江论职业教育 刘云山会客

尼泊尔地震致3904人遇难 首都因地震南移3米

4名在尼中国公民遇难 中国空军4架伊尔-76运输机赴尼泊尔救灾

“妖股”暴风科技：股价已透支1到2年业绩

[观点]虚拟现实业务不过是IPO讲故事 即使市场爆发 暴风也无任何内容优

百度资助国际救援组织赴尼泊尔救灾

- 综述：尼泊尔滞留华人就地“转型”做志愿者
- 尼泊尔貌拒台湾派搜救队 台：希望不要做政治联想
- 美日修改防卫合作指针 克里声称合作涵盖钓鱼岛
- 内地5月起禁止发布“民间天气预报”最高罚款5万
- 传央企将重组减至40家 国资委否认

“滴滴打车”遭索8000万或成闹剧 商标保护确有硬伤

- 绕不开物业又无法深入 社区O2O成“笑話”
- 科技大事件：微软暗示将放弃数十亿收购的诺基亚
- 韩总统朴槿惠批准韩国总理李完九辞职
- 交通部红十字总会等10部门三公经费不降反升(图)
- 山西巡视组：长治医学院个别领导实为学匪

中石化总经理王天普被查 知情人士称涉周滨案

- 广东副省长林少春任省政法委书记(图)
- 全国游艇会超70家多数被包装成公共项目拿地



智利艺术家将冰岛温泉染成粉红色

滚动新闻 中用气险恶 | 有人欲进藏救援 专业救援机构提醒：时机不成熟 | 欧莱雅生产环节二氧化氮

百度百家 热点 互联网 文化 娱乐 体育 财经 加入百家号



风投不上门 移动电源厂商大死亡

- 地图是未来科技世界最重要的配角
- 行业人士评Apple Watch：像个玩具
- 虚拟现实难以支撑暴风科技百亿市值
- 知乎的自我矛盾：如何搞定版权

康斯坦丁

罗超

国仁

一百榆尽

李书航



多方混战 洗类O2O前途可期

- 互联网团队管理的五项理论和实战
- 创业冷思考：大部分上门服务不成立
- A轮融资变难 给创业者的8条建议

刘旷

董飞

杨子超

硅发布

Interactive recommendation

- Number of recommendations k to choose from large
 - Similar articles → similar click-through rates!
- Performance depends on query / context
 - Similar users → similar click-through rates!
- Need to compile sets of k recommendations. (instead of only one)
 - Similar sets → similar click-through rates!

News recommendation

News

U.S. edition Modern

+Zaiwen

Personalize

Top Stories

National Guard Called Out in Baltimore as Police and Youths Clash After ...
New York Times - 13 minutes ago

BALTIMORE - Rioters in northwest Baltimore looted stores and pelted riot-gear-clad police with rocks on Monday, hours after Freddie Gray, the 25-year-old black man who has become the nation's latest symbol of police brutality, was laid to rest amid ...

Latest on police-custody death: Governor declares emergency Miami Herald
Maryland governor declares state of emergency in Baltimore amid rioting Los Angeles Times

Trending on Google+: Baltimore protests turn violent; police officers attacked CNN
In Depth: Maryland Gov. activates National Guard as Baltimore protests rage New York Daily News
Wikipedia: Death of Freddie Gray

Related Baltimore » Death »

See realtime coverage

Sorrow prevails Nepal capital after deadly quake
Xinhua - 9 minutes ago

KATHMANDU, April 27 (Xinhua) -- Smoke from burning pyres for bodies rose high and spread wide. The Nepalese, who are a quiet people, restrained from crying out loud.

In US-Japan talks, China is the elephant in the room
San Francisco Chronicle - 2 hours ago

Japanese Foreign Minister Fumio Kishida, left, and Defense Minister Gen Nakatani, second from left, attend a meeting with U.S. Secretary of State John Kerry, third from right, and Secretary of Defense Ashton Carter, not visible, in New York, Monday, April 27, ...

Nepal earthquake: RAF plane leaves for Nepal with UK aid
BBC News - 45 minutes ago

An RAF plane carrying UK aid supplies and a team of British Army Gurkha engineers is on its way to Nepal. Dozens of British and Irish people have still not been traced following Saturday's devastating earthquake.

Apple Earnings Surge 33% on iPhone Sales
Wall Street Journal - 11 minutes ago

Apple Inc. AAPL 1.82 % is pulling off a feat rarely seen in any industry, much less the cutthroat world of

Get Google News on the go.
Try the free app for your phone or tablet.

Recent

World 'Closer Than Ever' to Iran Nuclear Deal, Kerry Says
ABC News - 6 minutes ago

Here's the Most Surprising Thing About Apple's Crazy Earnings
TIME - 14 minutes ago

Josh Hamilton on Arte Moreno: 'He knew what the deal was'
USA TODAY - 12 minutes ago

Weather for Oakland, California

Today	Tue	Wed	Thu
73° 51°	71° 49°	72° 51°	83° 59°

The Weather Channel - Weather Underground - AccuWeather

Sports scores

Today	Yesterday
NHL	
NYI 0 • 0 11:47 1P	WAS
TB 2 • 0 0:00 1P	DET

MLB

connecting to ssl.gstatic.com...

Which set of articles satisfies most users?

News recommendation

Baidu 推荐

推荐 互联网 科技 财经 国际 军事 娱乐 体育 社会 国内 搜索



03 西藏吉隆镇至中尼边境热索桥道路中断

百度资助国际救援组织赴尼泊尔救灾

百度基金会通过联合国驻尼泊尔协调机构资助的2支国际救援行组织已经在尼泊尔地震区域展开救援行动，这也是...

412022 2221 407 分享

小牛主帅自信逆转淘汰火箭

视频：火箭109-121小牛、魔兽再冲突蒙塔31分爆哈登，时长约2分26秒。

7429 12 19 分享

何炅开心卖萌庆祝41岁生日

谢娜更新微博，晒出为何炅庆生的照片，并写道，录完节目吃生日蛋糕咯。

0 167 12 分享

戴相龙接受专访：退休后大部分时间和家人在一

登录 更懂你

下次自动登录 忘记密码？

立即注册

可以使用以下方式登录

微博 QQ 微信 腾讯微博

热门新闻排行榜

1	52岁蓝洁瑛再曝近照 烟不离手	9315
2	10086伪基站致银行卡信息外泄	9157
3	杨紫琼回忆尼泊尔惊魂一幕：马伊琍穿破洞裤现身变回潮妈	7778
4	二手苹果去哪儿了？80%流入	7729
5	京津冀一体化42万亿盛宴将启	6320
6	昆凌孕肚明显仍不知宝宝性别	5963
7	上海：车模扮乞丐抗议车展取乐	5898
8	山西巡视组：长治医学院个别领导	5730
9	二手苹果去哪儿了？80%流入	5640

热门科技新闻排行榜

1	10086伪基站致银行卡信息外泄	7267
2	二手苹果去哪儿了？80%流入	6039

意见反馈

Which set of articles satisfies most users?

Sponsored search

Google search results for "cell phone":

About 1,180,000,000 results (0.68 seconds)

Organic Results:

- iPhone 6**
www.apple.com - Bigger than bigger. Learn more.
9 5656 Bay St, Emeryville, CA - (510) 350-2400
Buy now Design iOS 8
- The Samsung Galaxy S®6 - Samsung.com**
www.samsung.com/GALAXY - Discover The New, Innovative Samsung Galaxy S®6 - Available Now. "The Galaxy S6 is Samsung's best smartphone ever." - Mashable Ratings: Prices 10/10 - Features 10/10 - Overall 10/10 - Reliability 9/10 Official YouTube - Follow Us on Twitter - Watch GS®6 Unboxing Video
- Consumer Cellular® - Our Low Rates Fit Your Needs**
www.consumercellular.com - Plans Start at \$10. Find Out More. No Contracts • \$10/Month Plans • 30-Day Guarantee • 100% Risk Free Consumer Cellular has 1,605 followers on Google+ AARP Discounts - Text & Data Plans - Phones - Voice Plans
- Mobile phone - Wikipedia, the free encyclopedia**
en.wikipedia.org/wiki/Mobile_phone - Wikipedia - A mobile phone (also known as a cellular phone, cell phone, hand phone, or simply a phone) is a phone that can make and receive telephone calls over a radio ... History of mobile phones - Cell Phone (film) - DynaTAC - Martin Cooper
- Smartphones & Cell Phones | Compare our Best ... - T-Mobile**
www.t-mobile.com/cell-phones.html - T-Mobile US, Inc. - Shop at T-Mobile and compare our best selection of cell phones and ... IF YOU CANCEL WIRELESS SERVICE, REMAINING BALANCE ON PHONE BECOMES ...
- Cell Phones - Mobile Phones - Walmart.com**
www.walmart.com/cp/mobile-phones/1105910 - Walmart - Looking for new cell phones? Shop for new cell phones, iPhones, unlocked phones, iPhone accessories, contract mobile phones and more Walmart.com.

Sponsored Ads:

Shop for cell phone on Google

Samsung Galaxy Note Edge (with con... \$0.00 Sprint	Apple iPhone 5s 16GB (with contract) \$0.00 Sprint	Samsung - Galaxy S 4 4g Cell Phone - W... \$1.00 Best Buy In store
Samsung Galaxy S 6 edge 64GB (with co... \$0.00 Sprint	Samsung - Galaxy S 5 4g LTE Cell Phone... \$1.00 Best Buy In store	Verizon LG Optimus Zone 2 Prepaid Sim... \$29.88 Walmart

Map for cell phone

Pre-Owned Cell Phones
buy.gazelle.com/Used-Cell-Phones - No Contracts & 30-Day Guarantee. Buy a Certified Like New Phone!

Cheapest Cell Phones
www.smartprices.us/Cell-Phones - Top Brands. Latest Models. Compare & Save Big. All Top Providers

Connecting to www.ostatic.com... 10:32AM 1 sketchre

Which set of ads should be displayed to maximize revenue?

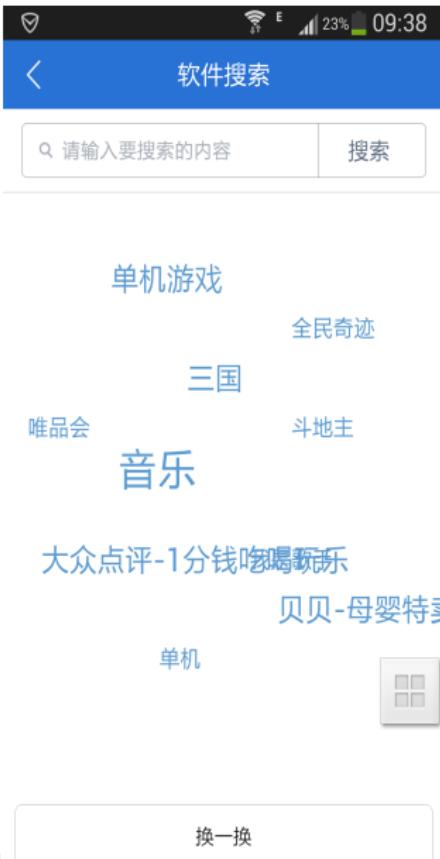


Sponsored search

Which set of ads should be displayed to maximize revenue?

Relevance vs. Diversity

- Users may have different interests / queries may be ambiguous
 - E.g., "jaguar", "squash", ...
- Want to choose a set that is relevant to as many users as possible
 - Users may choose from the set the article they're most interested in
- Want to optimize both relevance and diversity



Simple abstract model

- Suppose we're given a set W of users and a collection V of articles/ads
- Each article i is relevant to a set of users S_i
 - For now suppose this is known!
- For each set of articles define

$$F(A) = |\cup_{i \in A} S_i|$$

- Want to select k articles to maximize "users covered"

$$\max_{|A| < k} F(A)$$

- Number of sets A grows exponential in k !
- Finding optimal A is NP-hard

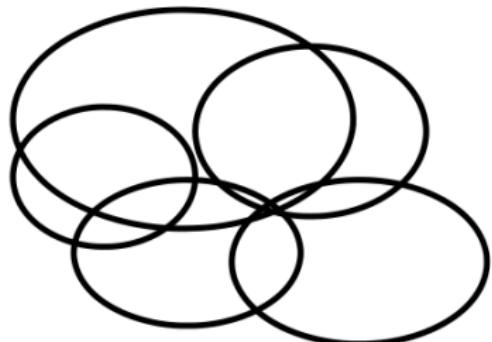
Maximum coverage

- Given: Collection V of sets, utility function $F(A)$

Want: $A^* \subseteq V$ such that

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

NP-hard!



Greedy algorithm:

Start with $A_0 = \{\}$

For $i=1$ to k

$$s^* = \operatorname{argmax}_s F(A_{i-1} \cup \{s\})$$

$$A_i = A_{i-1} \cup \{s^*\}$$

How well does
this simple
heuristic do?

Approximation guarantee

Theorem

Under some natural conditions, greedy algorithm produces a solution A, where $F(A) \geq (1 - 1/e) * \text{optimal-value} (\sim 63\%)$.

[Nemhauser, Fisher, Wolsey '78]

- This result holds for utility functions F with 2 properties:
 - F is **(nonnegative) monotone**:
if $A \subseteq B$ then $0 \leq F(A) \leq F(B)$
 - F is **submodular**:
"diminishing returns"

Outline

1 What is submodularity?

- Examples in recommendation sets
- Definition

2 Submodular maximization

3 Submodular minimization

4 Applications of submodular maximization

Set Functions

- **Ground set:** subsets of some finite set
- Given a set X , the set $V := 2^X = \{A \mid A \subseteq X\}$
- A **set function** takes as input a set, and outputs a real number
 - Inputs are subsets of some **ground set** X
 - $F : 2^X \rightarrow \mathbb{R}$
- It is common in the literature to use either X or V as the ground set.
- We will follow this inconsistency in the literature and will inconsistently use either X or V as our ground set (hopefully not in the same equation, if so, please point this out).

Set Functions

- If F is a modular function, then for any $A, B \subseteq V$, we have

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

- If F is a modular function, it may be written as

$$F(A) = F(\emptyset) + \sum_{a \in A} (F(\{a\}) - F(\emptyset))$$

- **modular set functions**
 - Associate a weight w_i with each $i \in X$, and set $F(S) = \sum_{i \in S} w_i$
 - Discrete analogue of linear functions
- Other possibly useful properties a set function may have:
 - **Monotone**: if $A \subseteq B \subseteq X$, then $F(A) \leq F(B)$
 - **Nonnegative**: $F(A) \geq 0$ for all $S \subseteq X$
 - **Normalized**: $F(\emptyset) = 0$.

Submodular Functions

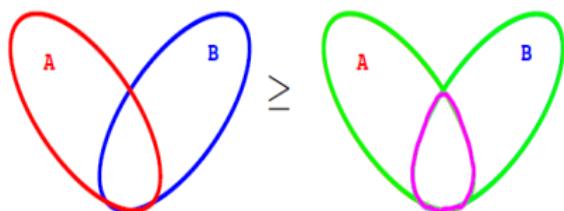
Definition 1

A set function $F : V \rightarrow \mathbb{R}$ is **submodular** if and only if

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

for all $A, B \subseteq V$.

- “Uncrossing” two sets reduces their total function value



Definition

A set function $F : V \rightarrow \mathbb{R}$ is **supmodular** if and only if $-F$ is submodular.

Submodular Functions

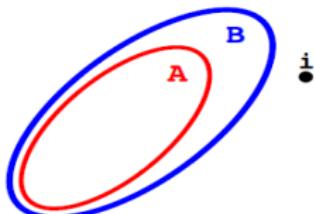
Definition 2 (diminishing returns)

A set function $F : V \rightarrow \mathbb{R}$ is **submodular** if and only if

$$\underbrace{F(B \cup \{s\}) - F(B)}_{\text{Gain of adding a set } s \text{ to a large solution}} \leq \underbrace{F(A \cup \{s\}) - F(A)}_{\text{Gain of adding a set } s \text{ to a small solution}}$$

for all $A \subseteq B \subseteq V$ and $s \notin B$.

- The marginal value of an additional element exhibits “diminishing marginal returns”
- This means that the incremental “value”, “gain”, or “cost” of s decreases (diminishes) as the context in which s is considered grows from A to B .



Submodular: Consumer Costs of Living

- Consumer costs are very often submodular. For example:

$$f(\text{Hamburger} + \text{French Fries}) + f(\text{Beverage}) \geq f(\text{Hamburger} + \text{Beverage}) + f(\text{French Fries})$$

- When seen as diminishing returns:

$$f(\text{French Fries} + \text{Beverage}) - f(\text{French Fries}) \geq f(\text{Hamburger} + \text{Beverage}) - f(\text{Hamburger})$$

Submodular Functions

Definition 3 (group diminishing returns)

A set function $F : V \rightarrow \mathbb{R}$ is **submodular** if and only if

$$F(B \cup C) - F(B) \leq F(A \cup C) - F(A)$$

for all $A \subseteq B \subseteq V$ and $C \subseteq V \setminus B$.

- This means that the incremental “value”, “gain”, or “cost” of set C decreases (diminishes) as the context in which C is considered grows from A to B .

Equivalence of Definitions

Definition 2 \implies Definition 3

Let $C = \{c_1, \dots, c_k\}$. The Definition 2 implies

$$\begin{aligned} & F(A \cup C) - F(A) \\ = & F(A \cup C) - \sum_{i=1}^{k-1} (F(A \cup \{c_1, \dots, c_i\}) - F(A \cup \{c_1, \dots, c_i\})) - F(A) \\ = & \sum_{i=1}^k (F(A \cup \{c_1, \dots, c_i\}) - F(A \cup \{c_1, \dots, c_{i-1}\})) \\ \geq & \sum_{i=1}^k (F(B \cup \{c_1, \dots, c_i\}) - F(B \cup \{c_1, \dots, c_{i-1}\})) \\ = & F(B \cup C) - F(B) \end{aligned}$$

Equivalence of Definitions

Definition 1 \implies Definition 2

To prove (2), let $A' = A \cup \{i\}$ and $B' = B$ and apply (1)

$$\begin{aligned} F(A \cup \{i\}) + F(B) &= F(A') + F(B') \\ &\geq F(A' \cap B') + F(A' \cup B') \\ &= F(A) + F(B \cup \{i\}) \end{aligned}$$

Definition 2 \implies Definition 1

Assume $A \neq B$. Define $A' = A \cap B$, $C = A \setminus B$ and $B' = B$. Then

$$\begin{aligned} F(A' \cup C) - F(A') &\geq F(B' \cup C) - F(B') \\ \iff F((A \cap B) \cup (A \setminus B)) + F(B) &\geq F(B \cup (A \setminus B)) + F(A') \\ \iff F(A) + F(B) &\geq F(A \cup B) + F(A \cap B) \end{aligned}$$

Submodularity

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

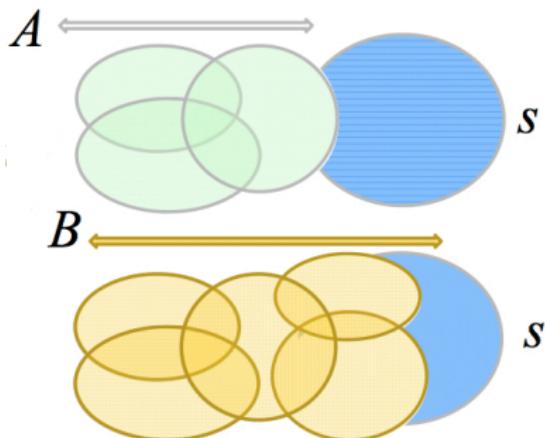
Example: Set cover

- F is submodular: $A \subseteq B$

$$\underbrace{F(A \cup \{s\}) - F(A)}_{\text{Gain of adding a set } s \text{ to a small solution}} \geq \underbrace{F(B \cup \{s\}) - F(B)}_{\text{Gain of adding a set } s \text{ to a large solution}}$$

- Natural example:
 - Set S_1, S_2, \dots, S_n
 - $F(A) = \text{size of union of } S_i$
(e.g., "number of satisfied users")

$$F(A) = |\cup_{i \in A} S_i|$$



Closedness properties

- F_1, \dots, F_m submodular functions on V and $\lambda_1, \dots, \lambda_m \geq 0$
 - 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 $\Rightarrow \sum_\theta P(\theta)F_\theta(A)$ submodular!
 - Multi-objective optimization:
 F_1, \dots, F_m submodular, $\lambda_i > 0 \Rightarrow \sum_i \lambda_i F_i(A)$ submodular

Probabilistic set cover

- Document coverage function:

$\text{cover}_d(c)$ =probability document d covers concept c , e.g., how strongly d covers c

It can model how relevant is concept c for user u

- Set coverage function:

$$\text{cover}_A(c) = 1 - \prod_{d \in A} (1 - \text{cover}_d(c))$$

Probability that at least one document in A covers c

- Objective:

$$\max_{|A| \leq k} F(A) = \sum_c w_c \text{cover}_A(c)$$

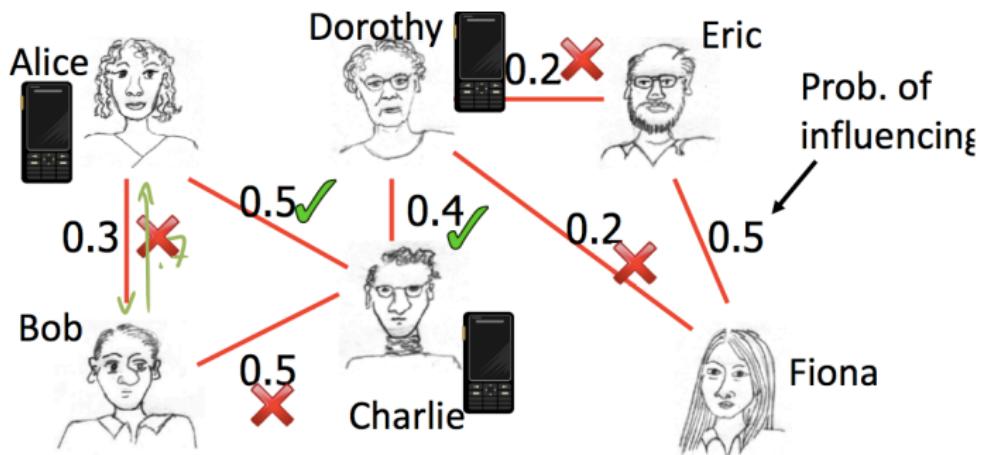
w_c is the concept weights

- The objective function is submodular

A model of Influence in Social Networks

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

Example: Influence in social networks [Kempe, Kleinberg, Tardos KDD'03]

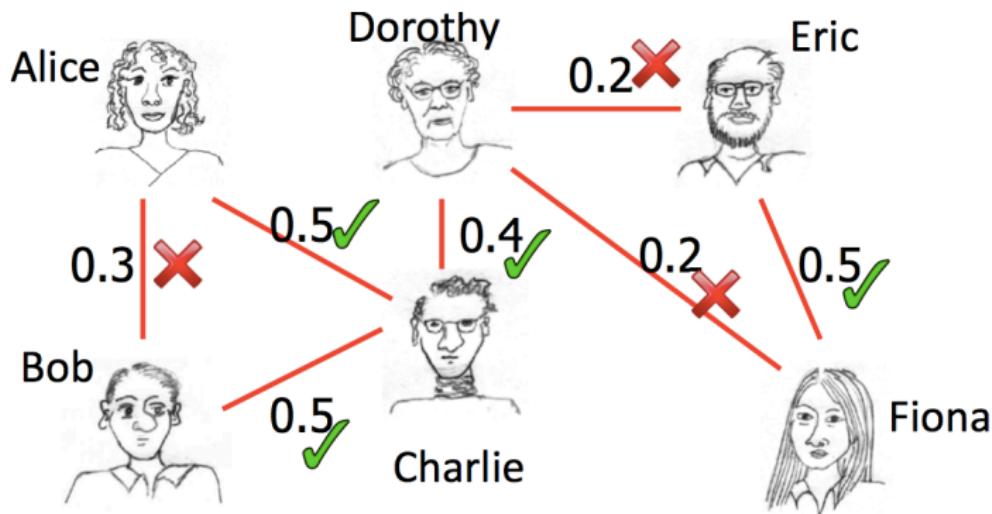


Which nodes are most influential?

$V = \text{Alice, Bob, Charlie, Dorothy, Eric, Fiona}$

$F(A) = \text{Expected number of people influenced by set } A$

Influence in social networks is submodular [Kempe, Kleinberg, Tardos KDD'03]



Key idea: Flip coins c in advance \rightarrow "live" edges

$F_c(A) =$ People influenced under outcome c

$F(A) = \sum_c P(c)F_c(A)$ is submodular as well!

The value of a friend

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

Information and Summarization

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

Restriction

Restriction

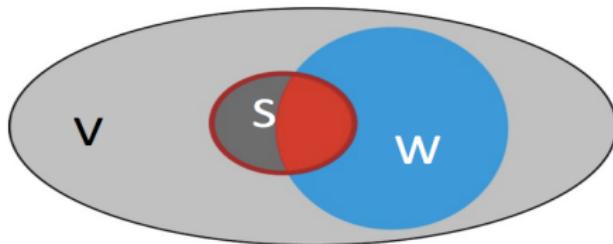
If $F(S)$ is submodular on V and $W \subseteq V$. Then $F'(S) = F(S \cap W)$ is submodular

Proof: Given $A \subseteq B \subseteq V \setminus \{i\}$, prove:

$$F((A \cup \{i\}) \cap W) - F(A \cap W) \geq F((B \cup \{i\}) \cap W) - F(B \cap W).$$

If $i \notin W$, then both differences on each size are zero. Suppose that $i \in W$, then $(A \cup \{i\}) \cap W = (A \cap W) \cup \{i\}$ and $(B \cup \{i\}) \cap W = (B \cap W) \cup \{i\}$. We have $A \cap W \subseteq B \cap W$, the submodularity of F yields

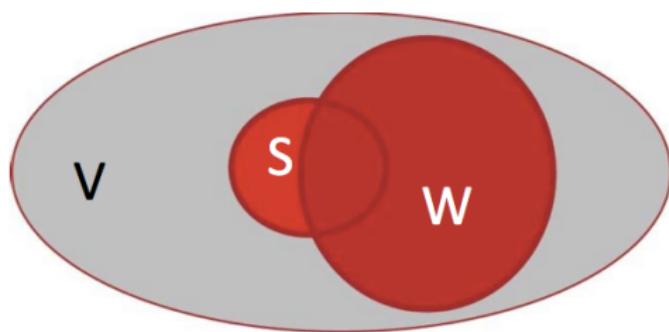
$$F((A \cap W) \cup \{i\}) - F(A \cap W) \geq F((B \cap W) \cup \{i\}) - F(B \cap W).$$



Conditioning

Conditioning

If $F(S)$ is submodular on V and $W \subseteq V$. Then $F'(S) = F(S \cup W)$ is submodular



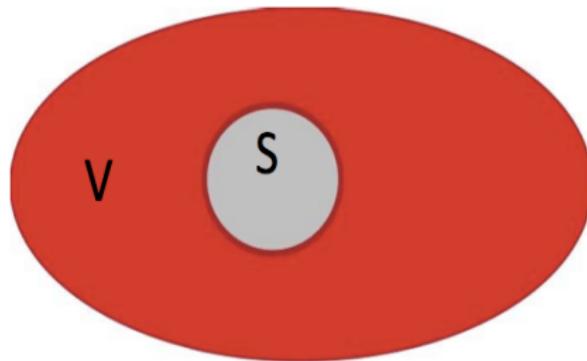
Reflection

Reflection

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

Proof: Since $V \setminus (A \cup B) = (V \setminus A) \cap (V \setminus B)$ and
 $V \setminus (A \cap B) = (V \setminus A) \cup (V \setminus B)$, then

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

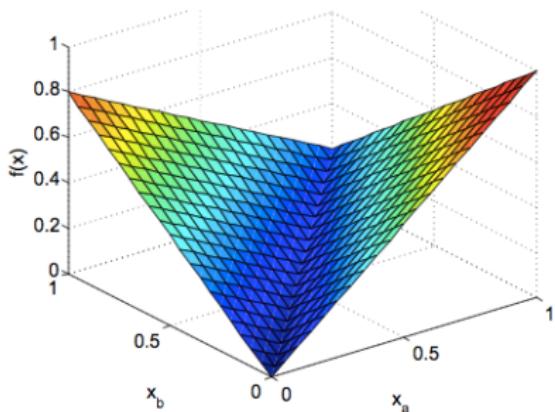


Convex aspects

- Submodularity as discrete analogue of convexity

- Convex extension
- Duality
- Polynomial time minimization!

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



- Many applications (computer vision, ML, ...)

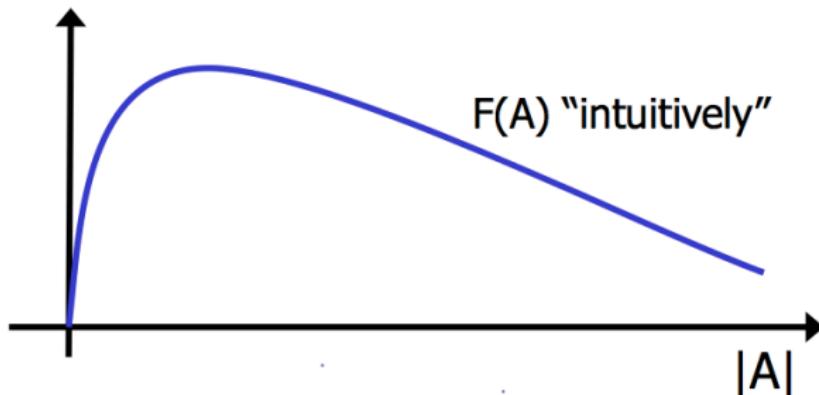
Concave aspects

- Marginal gain $\triangle_F(s|A) = F(\{s\} \cup A) - F(A)$
- Submodular:

$$\forall A \subseteq B, s \notin B : \quad F(A \cup \{s\}) - F(A) \geq F(B \cup \{s\}) - F(B)$$

- Concave:

$$\forall a \leq b, s > 0 \quad g(a + s) - g(a) \geq g(b + s) - g(b)$$



$$\forall a \leq b, s > 0 \quad g(a+s) - g(a) \geq g(b+s) - g(b)$$

- Suppose that $a+s \in [a, b]$
- Apply the concavity of $g(x)$ to $[a, a+s, b+s]$:

$$\begin{aligned} g(a+s) &\geq \frac{b-a}{b+s-a}g(a) + \frac{s}{b+s-a}g(b+s) \\ \iff g(a+s) - g(a) &\geq \frac{-s}{b+s-a}g(a) + \frac{s}{b+s-a}g(b+s) \end{aligned}$$

- Apply the concavity of $g(x)$ to $[a+s, b, b+s]$:

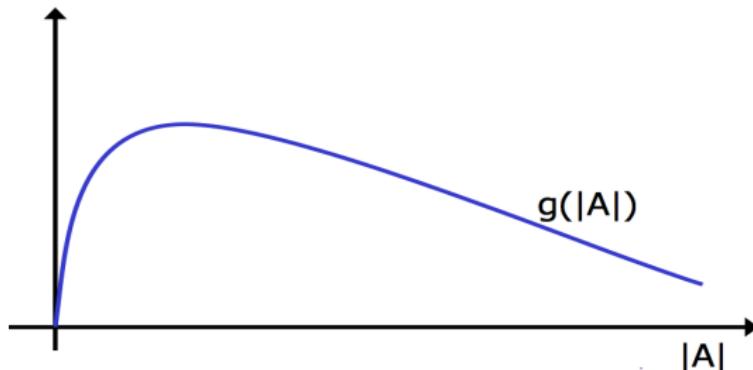
$$\begin{aligned} g(b) &\geq \frac{s}{b+s-a}g(a) + \frac{b-a}{b+s-a}g(b+s) \\ \iff g(b+s) - g(b) &\leq \frac{-s}{b+s-a}g(a) + \frac{s}{b+s-a}g(b+s) \end{aligned}$$

Submodularity and Concavity

Let $m \in \mathbb{R}_+^X$ be a modular function, and g a concave function over \mathbb{R} . Define $F(A) = g(m(A))$. Then $F(A)$ is submodular.

Proof: Given $A \subseteq B \subseteq X \setminus v$, we have $0 \leq a = m(A) \leq b = m(B)$, and $0 \leq s = m(v)$. For g concave, we have
$$g(a + s) - g(a) \geq g(b + s) - g(b), \text{ which implies}$$

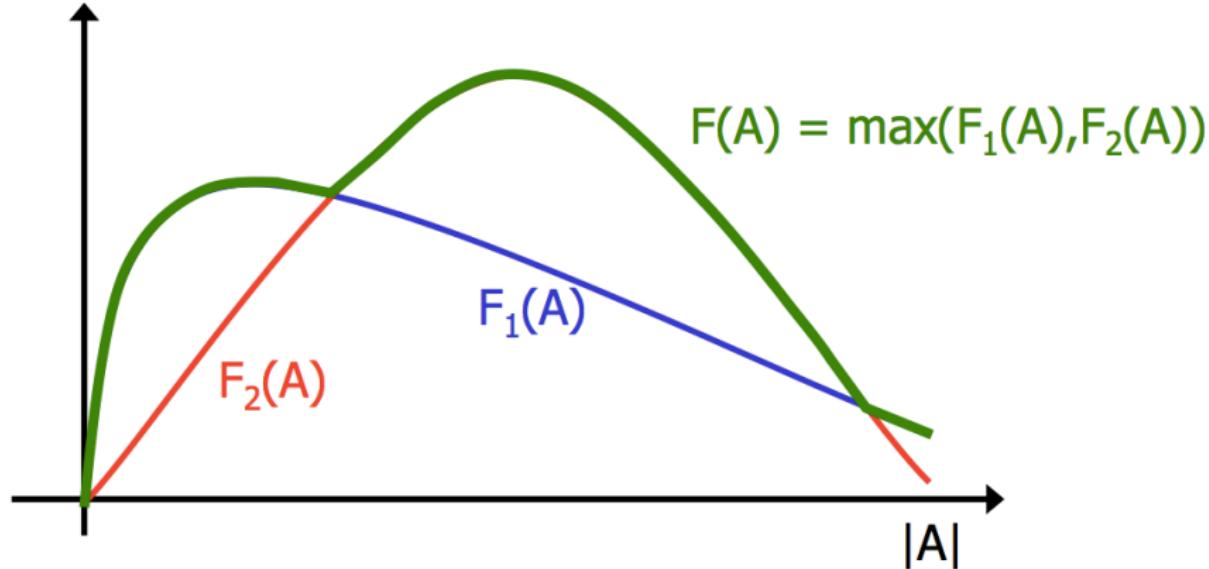
$$g(m(A) + m(v)) - g(m(A)) \geq g(m(B) + m(v)) - g(m(B))$$



Maximum of submodular functions

Suppose $F_1(A)$ and $F_2(A)$ submodular.

Is $F(A) = \max(F_1(A), F_2(A))$ submodular?



$\max(F_1, F_2)$ 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

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

<

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

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

Max - normalized

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

$$F(A) = \max_{j \in A} c_j$$

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

Proof: Since

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

and

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

we have

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

Monotone difference of two functions

Let F and G both be submodular functions on subsets of V and let $(F - G)(\cdot)$ be either monotone increasing. Then $h : 2^V \rightarrow \mathbb{R}$ defined by $h(A) = \min(F(A), G(A))$ is submodular.

- If $h(A)$ agrees with either f or g on both X and Y , the result follows since

$$\begin{aligned} F(X) + F(Y) \\ G(X) + G(Y) \end{aligned} \geq \min(F(X \cup Y), G(X \cup Y)) + \min(F(X \cap Y), G(X \cap Y))$$

- otherwise, w.l.o.g., $h(X) = F(X)$ and $h(Y) = G(Y)$, giving

$$h(X) + h(Y) = F(X) + G(Y) \geq F(X \cup Y) + F(X \cap Y) + G(Y) - F(Y)$$

Assume $F - G$ is monotonic increasing. Hence,
 $F(X \cup Y) + G(Y) - F(Y) \geq G(X \cup Y)$ giving

$$h(X) + h(Y) \geq G(X \cup Y) + F(X \cap Y) \geq h(X \cup Y) + h(X \cap Y)$$

- Let $F : 2^V \rightarrow \mathbb{R}$ be an increasing or decreasing submodular function and let k be a constant. Then the function $h : 2^V \rightarrow \mathbb{R}$ defined by

$$h(A) = \min(k; F(A))$$

is submodular

- In general, the minimum of two submodular functions is not submodular. However, when wishing to maximize two monotone non-decreasing submodular functions, we can define function $h : 2^V \rightarrow R$ as

$$h(A) = \frac{1}{2}(\min(k, F) + \min(k, G))$$

then h is submodular, and $h(A) \geq k$ if and only if both $F(A) \geq k$ and $G(A) \geq k$

Outline

1 What is submodularity?

- Examples in recommendation sets
- Definition

2 Submodular maximization

3 Submodular minimization

4 Applications of submodular maximization

Submodular maximization a Cardinality Constraint

Problem Definition

Given a **non-decreasing** and **normalized** submodular function $F : 2^X \rightarrow \mathbb{R}^+$ on a finite ground set X with $|X| = n$, and an integer $k \leq n$:

$$\max F(A), \text{ s.t. } |A| \leq k$$

Greedy Algorithm

- ▶ $A_0 \leftarrow \emptyset$, set $i = 0$
- ▶ While $|A_i| \leq k$
 - Choose $s \in X$ maximizing $F(A_i \cup \{s\})$
 - $A_{i+1} \leftarrow A_i \cup \{s\}$

Greedy maximization is near-optimal

Theorem[Nemhauser, Fisher& Wolsey'78]

For monotonic submodular functions, Greedy algorithm gives constant factor approximation

$$F(A_{\text{greedy}}) \geq \underbrace{(1 - 1/e)}_{\sim 63\%} F(A^*)$$

- Greedy algorithm gives **near-optimal** solution!
- For many submodular objectives: **Guarantees best possible** unless P=NP
- Can also handle more complex constraints

Contraction/Conditioning

Let $F : 2^X \rightarrow \mathbb{R}$ and $A \subseteq X$. Define $F_A(S) = F(A \cup S) - F(A)$.

Lemma: If F is monotone and submodular, then F_A is monotone, submodular, and normalized for any A .

- Proof: Monotone:

- Let $S \subseteq T$, then $F_A(S) = F(A \cup S) - F(A) \leq F(A \cup T) - F(A) = F_A(T)$

- Submodular. Let $S, T \subseteq X$:

$$\begin{aligned}F_A(S) + F_A(T) &= F(S \cup A) - F(A) + F(T \cup A) - F(A) \\&\geq F(S \cup T \cup A) - F(A) + F((S \cap T) \cup A) - F(A) \\&= F_A(S \cup T) + F_A(S \cap T)\end{aligned}$$

Lemma

If F is normalized and submodular, and $A \subseteq X$, then there is $j \in A$ such that $F(\{j\}) \geq \frac{1}{|A|}F(A)$

- Proof. If A_1 and A_2 partition A , i.e., $A = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$, then

$$F(A_1) + F(A_2) \geq F(A_1 \cup A_2) + F(A_1 \cap A_2) = F(A)$$

- Applying recursively, we get

$$\sum_{j \in A} F(\{j\}) \geq F(A)$$

- Therefore, $\max_{j \in A} F(\{j\}) \geq \frac{1}{|A|}F(A)$

Greedy maximization is near-optimal

Theorem[Nemhauser, Fisher & Wolsey'78]

For monotonic submodular functions, Greedy algorithm gives constant factor approximation

$$F(A_{\text{greedy}}) \geq (1 - 1/e)F(A^*)$$

- Proof: Let A_i be the working set in the algorithm
- Let A^* be optimal solution.
- We will show that the suboptimality $F(A^*) - F(A)$ shrinks by a factor of $(1 - 1/k)$ each iteration
- After k iterations, it has shrunk to $(1 - 1/k)^k \leq 1/e$ from its original value
- The algorithm choose $s \in X$ maximizing $F(A_i \cup \{s\})$. Hence:

$$F(A_{i+1}) = F(A_i) + F(A_i \cup \{s\}) - F(A_i) = F(A_i) + \max_j F_{A_i}(\{j\})$$

- By our lemmas, there is $j \in A^*$ s.t.

$$\begin{aligned}
 F_{A_i}(\{j\}) &\geq \frac{1}{|A^*|} F_{A_i}(A^*) \quad (\text{apply lemma to } F_{A_i}) \\
 &= \frac{1}{k} (F(A_i \cup A^*) - F(A_i)) \\
 &\geq \frac{1}{k} (F(A^*) - F(A_i))
 \end{aligned}$$

- Therefore

$$\begin{aligned}
 F(A^*) - F(A_{i+1}) &= F(A^*) - F(A_i) - \max_j F_{A_i}(\{j\}) \\
 &\leq \left(1 - \frac{1}{k}\right) (F(A^*) - F(A_i)) \\
 &\leq \left(1 - \frac{1}{k}\right)^k (F(A^*) - F(\emptyset))
 \end{aligned}$$

Scaling up the greedy algorithm [Minoux'78]

In round $i+1$,

- have picked $A_i = s_1, \dots, s_i$
- pick $s_{i+1} = \arg \max_s F(A_i \cup \{s\}) - F(A_i)$

i.e., maximize "marginal benefit" $\Delta(s|A_i)$

$$\Delta(s|A_i) = F(A_i \cup \{s\}) - F(A_i)$$

Key observation: Submodularity implies

$$\Delta(s | A_i) \geq \Delta(s | A_{i+1})$$

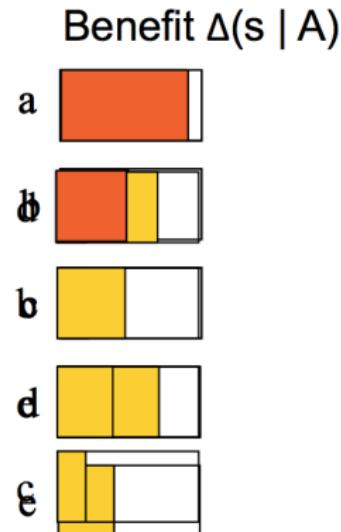


Marginal benefits can never increase!

"Lazy" greedy algorithm [Minoux'78]

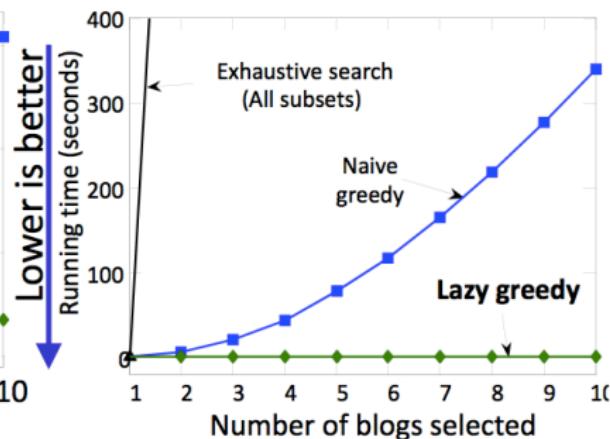
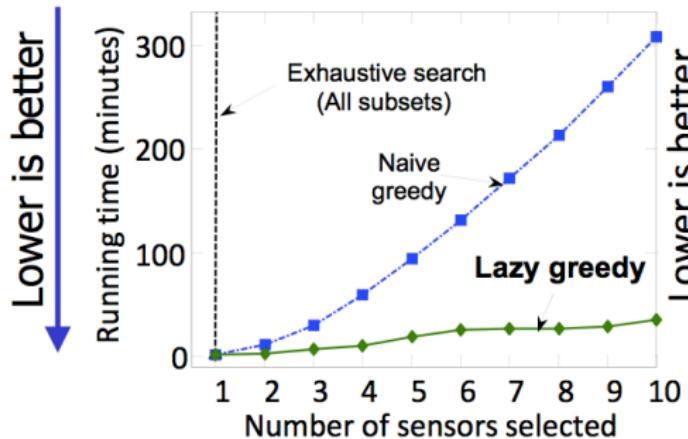
Lazy greedy algorithm:

- First iteration as usual
- Keep an ordered list of marginal benefits Δ_i from previous iteration
- Re-evaluate Δ_i only for top element
- If Δ_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



30x speedup

Blog selection



700x speedup

Outline

- 1 What is submodularity?
 - Examples in recommendation sets
 - Definition
- 2 Submodular maximization
- 3 Submodular minimization
- 4 Applications of submodular maximization

Optimizing Submodular Functions

- As our examples suggest, optimization problems involving submodular functions are very common
- These can be classified on two axes: constrained/unconstrained and maximization/minimization

	Maximization	Minimization
Unconstrained	NP-hard $\frac{1}{2}$ approximation	Polynomial time via convex opt
Constrained	Usually NP-hard $1 - 1/e$ (mono, matroid) O(1) ("nice" constraints)	Usually NP-hard to apx. Few easy special cases

Representation

In order to generalize all our examples, algorithmic results are often posed in the value oracle model. Namely, we only assume we have access to a subroutine evaluating $F(S)$.

Problem Definition

Given a submodular function $f : 2^X \rightarrow \mathbb{R}$ on a finite ground set X ,

$$\begin{aligned} & \min \quad F(S) \\ \text{s.t.} \quad & S \subseteq X \end{aligned}$$

- We denote $n = |X|$
- We assume $F(S)$ is a rational number with at most b bits
- Representation: in order to generalize all our examples, algorithmic results are often posed in the **value oracle** model. Namely, we only assume we have access to a subroutine evaluating $F(S)$ in constant time.

Goal

An algorithm which runs in time polynomial in n and b .

Some more notations

- $E = \{1, 2, \dots, n\}$
- $\mathbb{R}^E = \{x = (x_j \in \mathbb{R} : j \in E)\}$
- $\mathbb{R}_+^E = \{x = (x_j \in \mathbb{R} : j \in E) : x \geq 0\}$
- Any vector $x \in \mathbb{R}^E$ can be treated as a normalized modular function, and vice versa. That is

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

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

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

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

- given modular function $x \in \mathbb{R}^E$, we can write $x(A)$ in a variety of ways, i.e., $x(A) = x \cdot 1_A = \sum_{i \in A} x_i$

Continuous Extensions of a Set Function

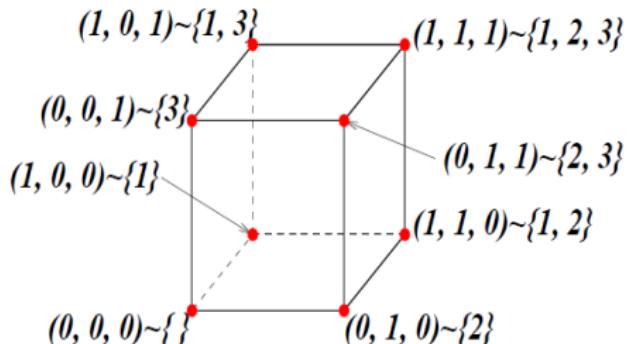
- A set function F on $X = \{1, \dots, n\}$ can be thought of as a map from the vertices $\{0, 1\}^n$ of the n -dimensional hypercube to the real numbers.

Extension of a Set Function

Given a set function $F : \{0, 1\}^n \rightarrow \mathbb{R}$, an extension of F to the hypercube $[0, 1]^n$ is a function $g : [0, 1]^n \rightarrow \mathbb{R}$ satisfying $g(x) = F(x)$ for every $x \in \{0, 1\}^n$.

$$\min_{w \in \{0,1\}^n} F(w)$$

with $\forall A \subseteq X, F(1_A) = F(A)$



Choquet integral - Lovász extension

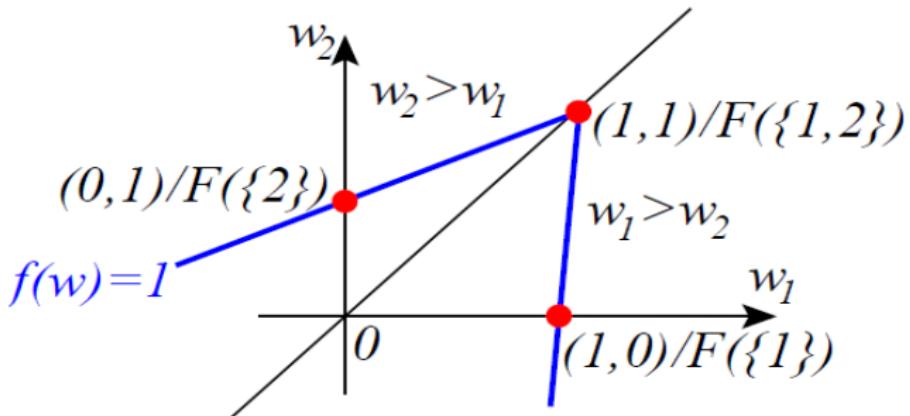
- Subsets may be identified with elements of $\{0, 1\}^n$
- Given **any** set-function F and w such that $w_{j_1} \geq \dots \geq w_{j_n}$, define

$$\begin{aligned} f(w) &= \sum_{k=1}^n w_{j_k} [F(\{j_1, \dots, j_k\}) - F(\{j_1, \dots, j_{k-1}\})] \\ &= \sum_{k=1}^{n-1} (w_{j_k} - w_{j_{k+1}}) F(\{j_1, \dots, j_k\}) + w_{j_n} F(\{j_1, \dots, j_n\}) \end{aligned}$$

- If $w = 1_A$, $f(w) = F(A) \implies$ extension from $\{0, 1\}^n$ to \mathbb{R}^n

Choquet integral - Lovász extension, example: $p = 2$

- If $w_1 \geq w_2$, $f(w) = F(\{1\})w_1 + [F(\{1, 2\}) - F(\{1\})]w_2$
- If $w_1 \leq w_2$, $f(w) = F(\{2\})w_2 + [F(\{1, 2\}) - F(\{2\})]w_1$



level set $\{w \in \mathbb{R}^2, f(w) = 1\}$ is displayed in blue

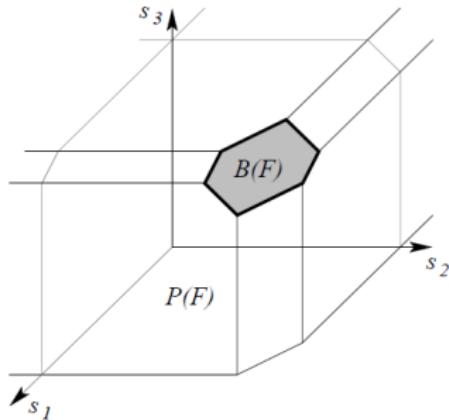
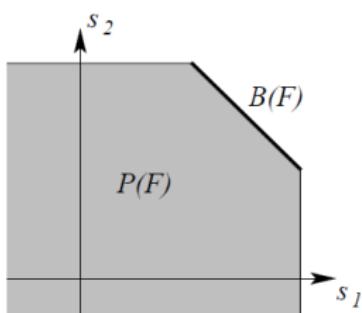
- Compact formulation: $f(w) = [F(\{1, 2\}) - F(\{1\}) - F(\{2\})] \min(w_1, w_2) + F(\{1\})w_1 + F(\{2\})w_2$

Links with convexity

Theorem (Lovász, 1982)

F is submodular if and only if f is convex

- Proof requires: Submodular and base polyhedra
- Submodular polyhedron: $P(F) = \{s \in \mathbb{R}^n, \forall A \subseteq V, s(A) \leq F(A)\}$
- Base polyhedron: $B(F) = P(F) \cap \{s(V) = F(V)\}$



Submodular and base polyhedra

- $P(F)$ has non-empty interior
- Many facets (up to 2^n), many extreme points (up to $n!$)

Fundamental property (Edmonds, 1970): If F is submodular, maximizing linear functions may be done by a “greedy algorithm”

- Let $w \in \mathbb{R}_+^n$ such that $w_{j_1} \geq \dots \geq w_{j_n}$
- Let $s_{j_k} = F(\{j_1, \dots, j_k\}) - F(\{j_1, \dots, j_{k-1}\})$ for $k \in \{1, \dots, n\}$
- Then

$$f(w) = \max_{s \in P(F)} w^\top s = \max_{s \in B(F)} w^\top s$$

- Both problems attained at s defined as above.
- proofs: pages 41-44 in http://bicmr.pku.edu.cn/~wenzw/bigdata/submodular_fbach_mlss2012.pdf

Links with convexity

Theorem (Lovász, 1982)

F is submodular if and only if f is convex

- If F is submodular, f is the maximum of linear functions. Then f is convex
- If f is convex, let $A, B \subseteq V$
 - $1_{A \cup B} + 1_{A \cap B} = 1_A + 1_B$ has components equal to 0 (on $V \setminus (A \cup B)$), 2 (on $A \cap B$) and 1 (on $A \Delta B = (A \setminus B) \cup (B \setminus A)$)
 - Thus $f(1_{A \cup B} + 1_{A \cap B}) = F(A \cup B) + F(A \cap B)$. Proof by writing out $f(1_{A \cup B} + 1_{A \cap B})$ and the definition of $f(w)$.
 - By homogeneity and convexity, $f(1_A + 1_B) \leq f(1_A) + f(1_B)$, which is equal to $F(A) + F(B)$, and thus F is submodular.

Links with convexity

Theorem (Lovász, 1982)

If F is submodular, then

$$\min_{A \subseteq V} F(A) = \min_{w \in \{0,1\}^n} f(w) = \min_{w \in [0,1]^n} f(w)$$

- Since f is an extension of F ,

$$\min_{A \subseteq V} F(A) = \min_{w \in \{0,1\}^n} f(w) \geq \min_{w \in [0,1]^n} f(w)$$

- Any $w \in [0, 1]^n$ can be decomposed as $w = \sum_{i=1}^m \lambda_i 1_{B_i}$, where $B_1 \subseteq \dots \subseteq B_m = V$, where $\lambda \geq 0$ and $\lambda(V) \leq 1$:
 - Since $\min_{A \subseteq V} F(A) \leq 0$ ($F(\emptyset) = 0$),

$$f(w) = \sum_{i=1}^m \lambda_i F(B_i) \geq \sum_{i=1}^m \lambda_i \min_{A \subseteq V} F(A) \geq \min_{A \subseteq V} F(A)$$

- Thus $\min_{w \in [0,1]^n} f(w) \geq \min_{A \subseteq V} F(A)$.

Links with convexity

- Any $w \in [0, 1]^n$, sort $w_{j_1} \geq \dots \geq w_{j_n}$. Find λ such that

$$\sum_{k=1}^n \lambda_{j_k} = w_{j_1}, \sum_{k=2}^n \lambda_{j_k} = w_{j_2}, \dots, \lambda_{j_n} = w_{j_n},$$

$$B_1 = \{j_1\}, B_2 = \{j_1, j_2\}, \dots, B_n = \{j_1, j_2, \dots, j_n\}$$

Then we have $w = \sum_{i=1}^n \lambda_i 1_{B_i}$, where $B_1 \subseteq \dots \subseteq B_n = V$, where $\lambda \geq 0$ and $\lambda(V) = \sum_{i \in V} \lambda_i \leq 1$.

Submodular function minimization

- Let $F : 2^V \rightarrow \mathbb{R}$ be a submodular function (such that $F(\emptyset) = 0$)
- convex duality:**

$$\begin{aligned}\min_{A \subseteq V} F(A) &= \min_{w \in [0,1]^n} f(w) \\ &= \min_{w \in [0,1]^n} \max_{s \in B(F)} w^\top s \\ &= \max_{s \in B(F)} \min_{w \in [0,1]^n} w^\top s = \max_{s \in B(F)} s_-(V)\end{aligned}$$

Submodular function minimization

Convex optimization

If F is submodular, then

$$\min_{A \subseteq V} F(A) = \min_{w \in \{0,1\}^n} f(w) = \min_{w \in [0,1]^n} f(w)$$

Using projected subgradient descent to minimize f on $[0, 1]^n$

- Iteration: $w_t = \Pi_{[0,1]^n}(w_{t-1} - \frac{C}{\sqrt{t}}s_t)$, where $s_t \in \partial f(w_{t-1})$
- $f(w) = \max_{s \in B(F)} w^\top s$
- Standard convergence results from convex optimization

$$f(w_t) - \min_{w \in [0,1]^n} f(w) \leq \frac{C}{\sqrt{t}}$$

Outline

- 1 What is submodularity?
 - Examples in recommendation sets
 - Definition
- 2 Submodular maximization
- 3 Submodular minimization
- 4 Applications of submodular maximization

Question

I have 10 minutes.Which blogs should I read to be most up to date?

[Leskovec, Krause, Guestrin, Faloutsos, VanBriesen, Glance'07]

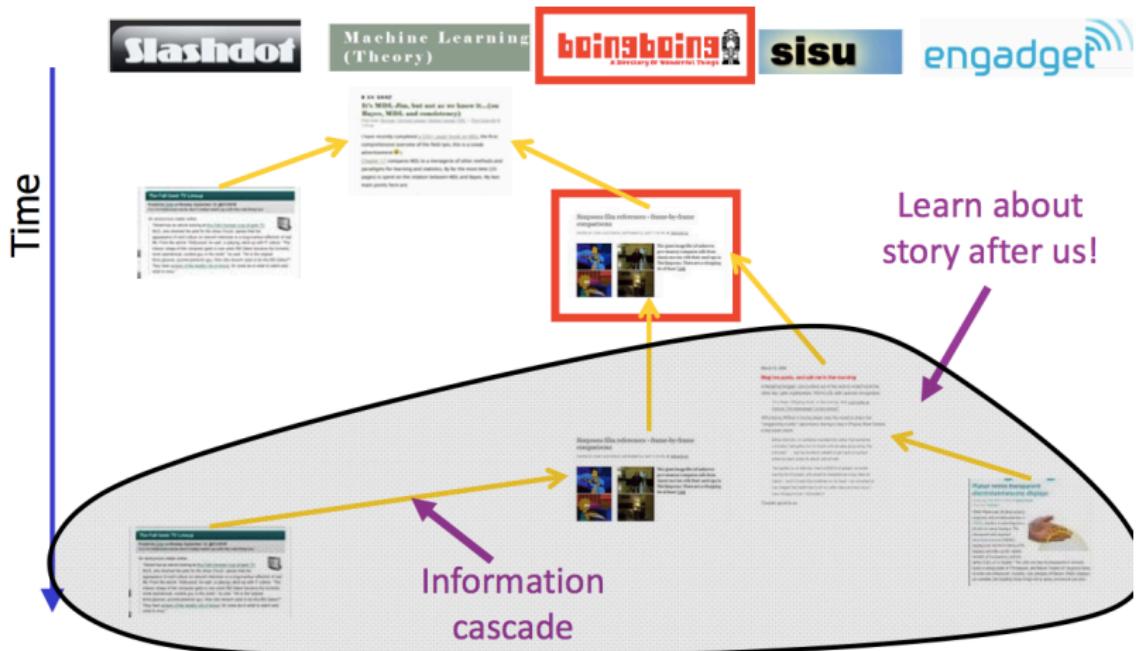


Thursday, Nov. 20, 2008

How Many Blogs Does the World Need?

By Michael Kinsley

Detecting Cascades in the Blogosphere



Which blogs should we read to learn about big cascades early?

Reward function is submodular

- Consider cascade i :

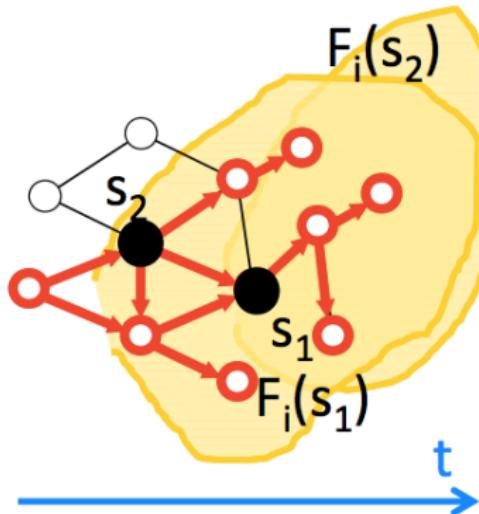
- $F_i(s_k)$ = benefit from blog s_k in event i
- $F_i(A) = \max F_i(s_k), s_k \in A$
 $\Rightarrow F_i$ is submodular

- Overall objective:

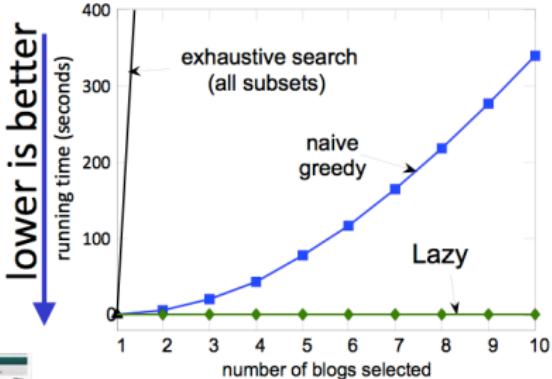
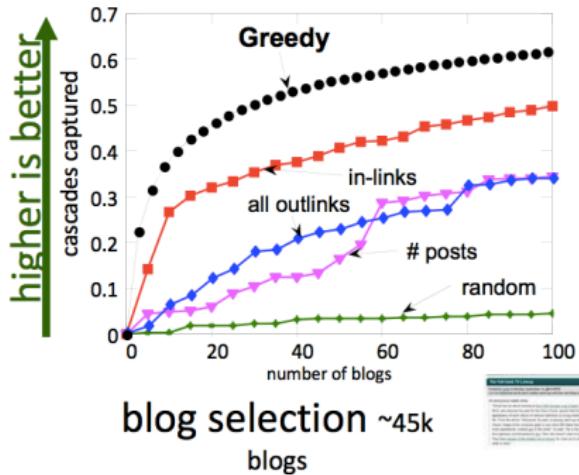
$$F(A) = \frac{1}{m} \sum_{i=1}^m F_i(A)$$

$\Rightarrow F$ is submodular

\Rightarrow Can use greedy algorithm to solve $\max_{|A| \leq k} F(A)$!

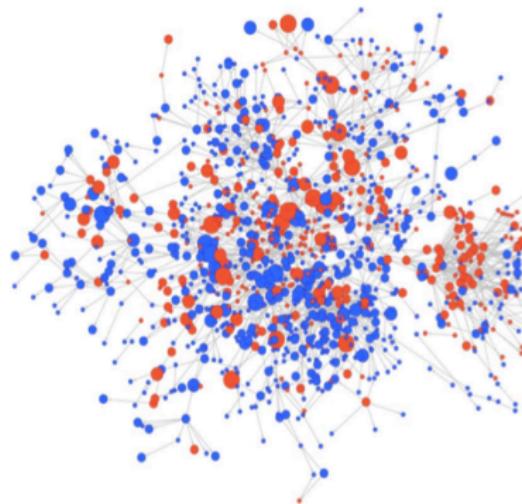
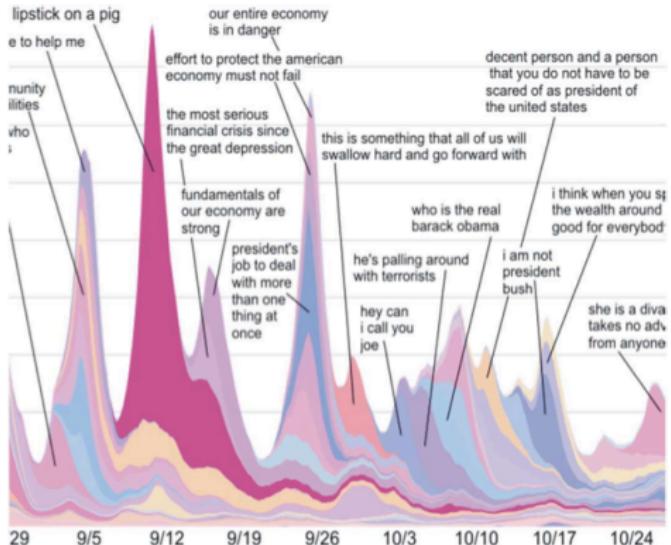


Performance on Blog selection



- Submodular formulation outperforms heuristics
- Lazy greedy gives 700x speedup

Application: Network inference



How can we learn who influences whom?

Inferring diffusion networks

[Gomez Rodriguez, Leskovec, Krause ACM TKDE 2012]

Given



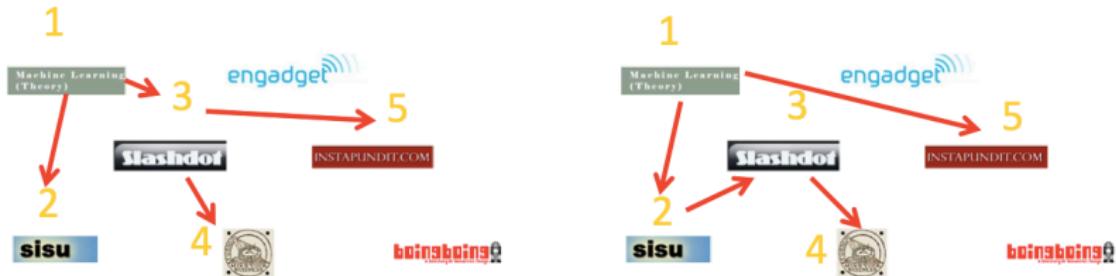
Want



Given traces of influence, wish to infer sparse directed network $G = (V, E)$
⇒ Formulate as optimization problem

$$E^* = \arg \max_{|E| \leq k} F(E)$$

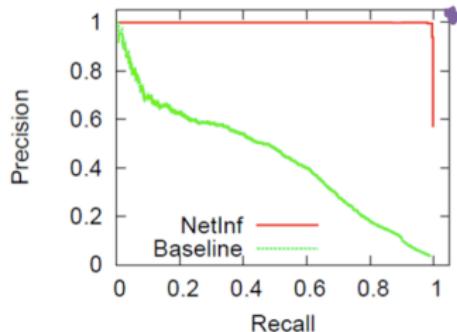
Estimation problem



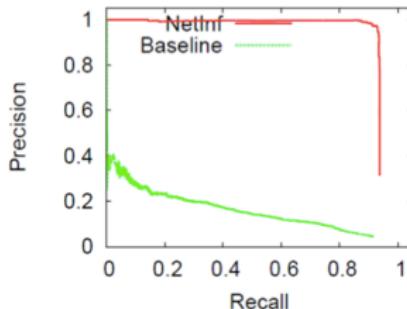
- Many influence trees T consistent with data
 - For cascade C_i , model $P(C_i|T)$
 - Find sparse graph that maximizes likelihood for all observed cascades
- ⇒ Log likelihood monotonic submodular in selected edges

$$F(E) = \sum_i \log \max_{\text{tree } T \subseteq E} P(C_i|T)$$

Evaluation: Synthetic networks



1024 node hierarchical Kronecker exponential transmission model

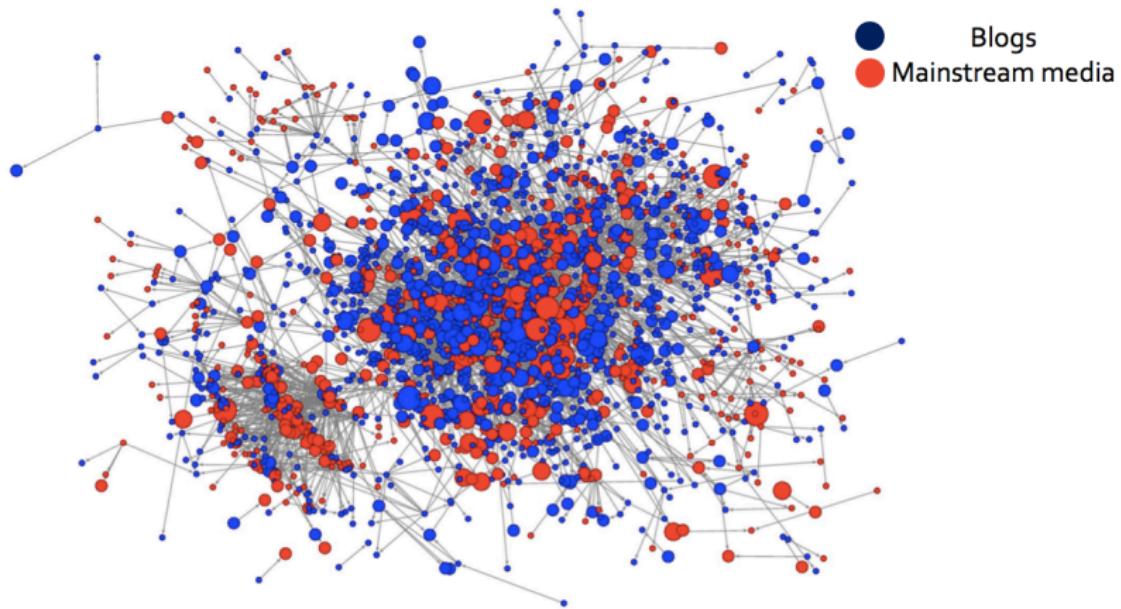


1000 node Forest Fire ($\alpha = 1.1$) power law transmission model

- Performance does not depend on the network structure:
 - Synthetic Networks: Forest Fire, Kronecker, etc.
 - Transmission time distribution: Exponential, Power Law
- Break-even point of > 90%

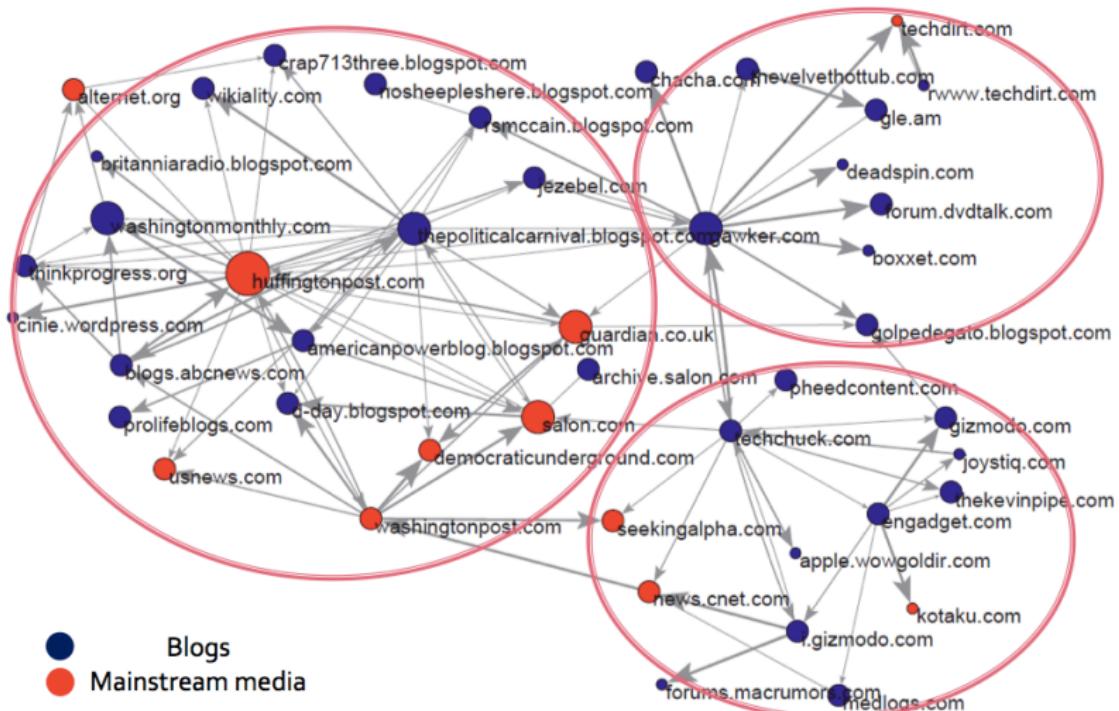
Inferred Diffusion Network

[Gomez Rodriguez, Leskovec, Krause ACM TKDE 2012]



Actual network inferred from 172 million articles from 1 million news sources

Diffusion Network (small part)



Application: Document summarization

[Lin & Bilmes'11]



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

Marginal gain of a sentence

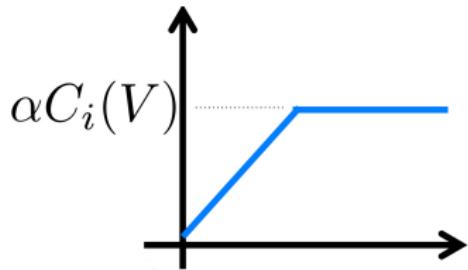


- Many natural notions of “document coverage” are submodular [Lin & Bilmes’11]

Relevance of a summary

$$F(S) = R(S) + \lambda D(S)$$

↑ ↑
Relevance Diversity



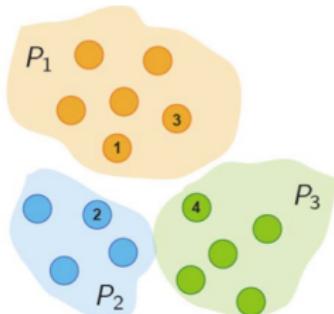
$$R(S) = \sum_i \min\{C_i(S), \alpha C_i(V)\}$$

$$C_i(S) = \sum_{j \in S} \omega_{i,j}$$

- $C_i(S)$: How well is sentence i "covered" by S
- $\omega_{i,j}$: Similarity between i and j

Diversity of a summary

$$D(S) = \sum_{i=1}^K \sqrt{\sum_{j \in P_i \cap S} r_j}$$
$$r_j = \frac{1}{N} \sum_i \omega_{i,j}$$



- r_j : Relevance of sentence j to doc.
- $\omega_{i,j}$: Similarity between i and j

Clustering of sentences
in document

Can be made query-specific; multi-resolution; etc.

Summary

- Many problems of recommending sets can be cast as submodular maximization
- Greedy algorithm gives best set of size k
- Can use lazy evaluations to speed up
- Approximate submodular maximization possible under a variety of constraints:
 - Matroid
 - Knapsack
 - Multiple matroid and knapsack constraints
 - Path constraints (Submodular orienteering)
 - Connectedness (Submodular Steiner)
 - Robustness (minimax)