Network Alignment, Graph Similarity, and Sports Ranking

David F. Gleich

Stanford University

Southeastern Ranking Workshop
College of Charleston

August 15th, 2009

In collaboration with
Mohsen Bayati, Margot Gerritsen, Amin Saberi, and Ying Wang.

Funded by Microsoft Live Labs Fellowship and the Library of Congress.

Administrivia

"Ladies and gentleman, please take your seat and fasten your seat belt. And also make sure your seat back and folding trays are in their full upright and locked position."

"At this time, we request that all cellular phones, pagers, radios and remote controlled toys be turned off for the full duration of the flight, as these items might interfere with the navigational and communication equipment on this aircraft."

PAPER `arxiv.org/abs/0907.3338`
Google "network alignment gleich"

CODE `stanford.edu/~dgleich/publications/2009/netalign`
Google "netalign gleich"

Desserts

URI: <http://id.loc.gov/authorities/sh85037243#concept>

Type: Topical Term

Broader Terms:

- Confectionery

Narrower Terms:

- Ambient desserts
- Banana splits
- Charlottes (Desserts)
- Chocolate desserts
- Frozen desserts
- Ice cream cones
- Mousses
- Puddings
- Refrigerated desserts
- Sundaes
- Whipped toppings

LC Classification: TX773

Created: 1986-02-11

Last Modified: 1988-01-15 17:36:44

Category:Desserts

From Wikipedia, the free encyclopedia

The main article for this category is ***dessert***.

Wikimedia Commons has media related to: ***Desserts***

Desserts are sweet foods eaten purely for pleasure, typically at the end of a meal.

Subcategories

This category has the following 16 subcategories, out of 16 total.

*****

- [+] Desserts by country (20)

B

- [+] Brand name desserts (3)

C

- [+] Cakes (0)
- [+] Chocolate desserts (0)
- [+] Confectionery (9)

C cont.

- [+] Cookies (1)
- [+] Custard desserts (0)

D

- [+] Dessert sauces (0)
- [+] Doughnuts (1)

F

- [+] Frozen desserts (2)

I

- [+] Ice cream (5)

P

- [+] Pastry (4)
- [+] Pies (9)
- [+] Puddings (2)

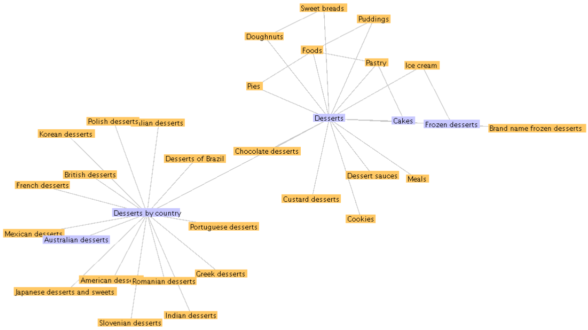
S

- [+] Sweet breads (2)

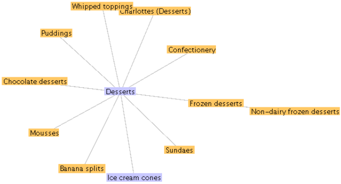
µ

- [+] Dessert stubs (1)

Library of Congress vs. Wikipedia



WC



LCSH

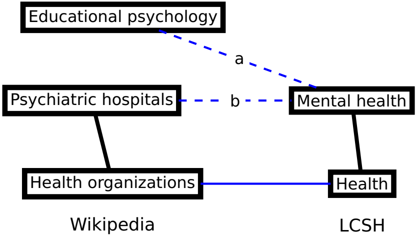
Are they similar?

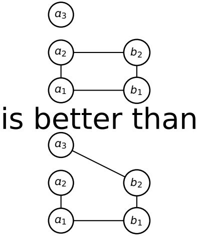
Note *Only a bit contrived, but I didn't have to work too hard to find the example*

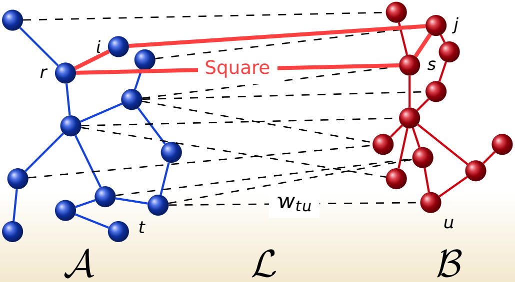
Network alignment

Slide 6 of 29

Alignment and overlap

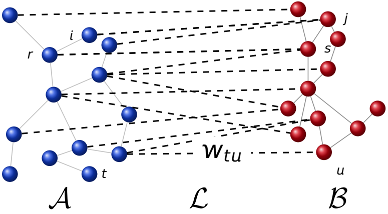






Maximize squares/overlap in 1-1 matching

LCSH/Wikipedia alignment



$\mathcal{A}$	LCSH	297,266 vertices, 248,230 edges
$\mathcal{B}$	Wikipedia	205,948 vertices, 382,353 edges
$\mathcal{L}$	links	4,971,629 edges

MAX-WEIGHT BIPARTITE MATCHING

Overlap	2,346
Weight	60,120 (106,294)

Using structure should do better!

Note *Links generated by a text search Lucene and weighted with Soft TF/IDF scoring*

Max-weight matching math

Variables, Data

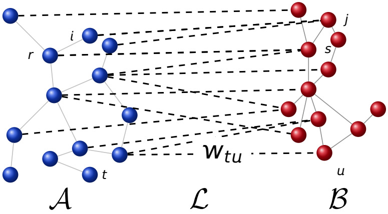
$x_i = $ edge indicator

$w_i = $ weight of edges

$e_i \in \mathcal{L}$

$e_i = (t, u)$

$w_i = w_{tu}$



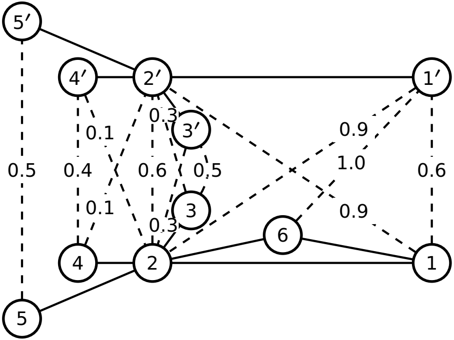
Problem

$$\begin{aligned} \underset{x_i}{\text{maximize}} \quad & \sum_{i:e_i \in \mathcal{L}} x_i w_i \\ \text{subject to} \quad & \sum_{i:e_i=(t,u)\in\mathcal{L}} x_i \le 1, \quad \forall t \in \mathcal{A} \\ & \sum_{i:e_i=(t,u)\in\mathcal{L}} x_i \le 1, \quad \forall u \in \mathcal{B} \\ & x_i \in \{0,1\} \end{aligned}$$

$\longleftrightarrow$

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbf{w}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{e} \\ & x_i \in \{0,1\} \end{aligned}$$

Example



$e_1 = (2, 2')$
$e_2 = (2, 1')$
$e_3 = (2, 3')$
$e_4 = (2, 4')$
$e_5 = (1, 2')$
$e_6 = (1, 1')$
$e_7 = (3, 2')$
$e_8 = (3, 3')$
$e_9 = (4, 2')$
$e_{10} = (4, 4')$
$e_{11} = (5, 5')$
$e_{12} = (6, 1')$

$$\mathbf{w} = \begin{pmatrix} 0.6 \\ 0.9 \\ 0.3 \\ 0.1 \\ 0.9 \\ 0.6 \\ 0.3 \\ 0.5 \\ 0.1 \\ 0.4 \\ 0.5 \\ 1.0 \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Matching and overlap

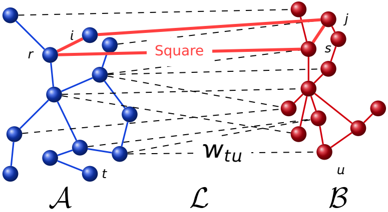
Squares produce overlap → bonus for some x_i *and* x_j → $\sum x_i x_j$

Variables, Data

$x_i =$ edge indicator $\qquad e_i \in \mathcal{L}$

$w_i =$ weight of edges $\qquad e_i = (t, u)$

S_{ij} squares in $\mathcal{S}$ $\qquad w_i = w_{tu}$



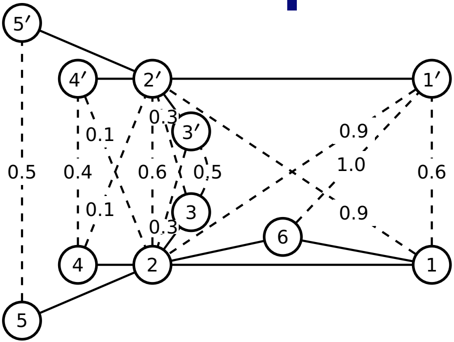
Problem

$$\underset{x_i}{\text{maximize}} \sum_{i:e_i \in \mathcal{L}} x_i w_i + \sum_{i,j \in \mathcal{S}} x_i x_j \quad \text{subject to} \quad x \text{ is a matching}$$

$\longleftrightarrow$

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbf{w}^T \mathbf{x} + \tfrac{1}{2} \mathbf{x}^T \mathbf{S} \mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{e} \\ & x_i \in \{0, 1\} \end{aligned}$$

Example with overlap

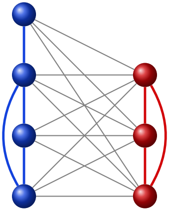


$e_1=(2,2')$
$e_2=(2,1')$
$e_3=(2,3')$
$e_4=(2,4')$
$e_5=(1,2')$
$e_6=(1,1')$
$e_7=(3,2')$
$e_8=(3,3')$
$e_9=(4,2')$
$e_{10}=(4,4')$
$e_{11}=(5,5')$
$e_{12}=(6,1')$

$$\mathbf{w} = \begin{pmatrix} 0.6 \\ 0.9 \\ 0.3 \\ 0.1 \\ 0.9 \\ 0.6 \\ 0.3 \\ 0.5 \\ 0.1 \\ 0.4 \\ 0.5 \\ 1.0 \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} 1&1&1&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&1 \\ 1&0&0&0&1&0&1&0&1&0&0&0 \\ 0&1&0&0&0&1&0&0&0&0&0&1 \\ 0&0&1&0&0&0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 0&0&0&0&0&1&0&1&0&1&1&1 \\ 0&0&0&0&1&0&1&0&1&0&0&0 \\ 0&0&0&0&1&0&1&0&1&0&0&0 \\ 0&0&0&0&1&0&1&0&1&0&0&0 \\ 0&1&1&1&0&0&0&0&0&0&0&1 \\ 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&1&1&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&1&1&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0&0&0&0&0 \\ 1&0&0&0&1&0&0&0&0&0&0&0 \end{pmatrix}$$

An important special case



When $\mathcal{L}$ contains everything possible *(the complete bipartite graph)*, then $\mathbf{S} = \mathbf{A} \otimes \mathbf{B}$ *(possibly permuted)*.

Important for discussion of related problems

Network alignment

NETWORK ALIGNMENT

$$\begin{array}{ll} \text{maximize} & \alpha \mathbf{w}^T \mathbf{x} + \frac{\beta}{2} \mathbf{x}^T \mathbf{S} \mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} \le \mathbf{e}, x_i \in \{0, 1\} \end{array}$$

History

- MAXIMUM COMMON SUBGRAPH
- PATTERN RECOGNITION
- ONTOLOGY MATCHING
- SIMILARITY FLOODING
- GRAPH SIMILARITY
- BIOINFORMATICS

Sparse problems

Sparse $\mathcal{L}$ largely ignored (exception Klau)
Our paper tackles that case explicitly

Conte el al. Thirty years of graph matching, 2004.; Melnik et al. Similarity flooding, 2004; Blondel et al. SIREV 2004; Singh et al. RECOMB 2007; Klau, BMC Bioinformatics 10:S59, 2009.

Algorithms

1. SNOPT Relax, solve QP with SNOPT
2. ISORANK *(skipped)*
3. LP Convert to LP, relax, solve
4. TIGHTLP Improve the LP
5. BP *(skipped)*
6. MR *(skipped)*

Large scale iterative methods skipped, see the paper.

Note *The* BP *algorithm is based on max-product belief propagation for a large-scale approach.* ISORANK *uses a PageRank inspired heuristic, and* MR *uses a sequence of matching problems as a sub-gradient approach.*

Singh et al. RECOMB2007; Klau, 2009

SNOPT

RELAXED NETWORK ALIGNMENT

$$\text{maximize} \quad \alpha \mathbf{w}^T \mathbf{x} + \frac{\beta}{2} \mathbf{x}^T \mathbf{S} \mathbf{x}$$
$$\text{subject to} \quad \mathbf{A}\mathbf{x} \le \mathbf{e}, 0 \le x_i \le 1$$

- **S** is indefinite *(still NP-hard)*
- only local maximum, no upper-bound on the solution

But x is still fractional! Use MAX-WEIGHT MATCHING with "weight" from the fractional **x** to "round" the solution.

SNOPT	
Overlap	11,155
Weight	49,236 (100,886)

Note Took 6 days to compute, did not reach an optimum.

LP

$$\mathbf{x}^T \mathbf{S} \mathbf{x} = \mathbf{e}^T \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & x_1x_6 & 0 & x_1x_8 & 0 & x_1x_{10} & x_1x_{11} & x_1x_{12} \\ 0 & 0 & 0 & 0 & x_2x_5 & 0 & x_2x_7 & 0 & x_2x_9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_3x_5 & 0 & x_3x_7 & 0 & x_3x_9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_4x_5 & 0 & x_4x_7 & 0 & x_4x_9 & 0 & 0 & 0 \\ 0 & x_5x_2 & x_5x_3 & x_5x_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_5x_{12} \\ x_6x_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_7x_2 & x_7x_3 & x_7x_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_8x_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_9x_2 & x_9x_3 & x_9x_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_{10}x_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_{11}x_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_{12}x_1 & 0 & 0 & 0 & x_{12}x_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{e}$$

$$\geq \mathbf{e}^T \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & y_{1,6} & 0 & y_{1,8} & 0 & y_{1,10} & y_{1,11} & y_{1,12} \\ 0 & 0 & 0 & 0 & y_{2,5} & 0 & y_{2,7} & 0 & y_{2,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{3,5} & 0 & y_{3,7} & 0 & y_{3,9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{4,5} & 0 & y_{4,7} & 0 & y_{4,9} & 0 & 0 & 0 \\ 0 & y_{5,2} & y_{5,3} & y_{5,4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{5,12} \\ y_{6,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{7,2} & y_{7,3} & y_{7,4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{8,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{9,2} & y_{9,3} & y_{9,4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{10,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{11,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_{12,1} & 0 & 0 & 0 & y_{12,5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{e}; \quad \begin{matrix} y_{i,j} \leq x_i \\ y_{i,j} \leq x_j \end{matrix}$$

LP

NETWORK ALIGNMENT LP

$$\begin{aligned} \underset{\mathbf{x},\mathbf{y}}{\text{maximize}} \quad & \alpha \mathbf{w}^T \mathbf{x} + \tfrac{\beta}{2} \mathbf{e}^T \mathbf{Y_S} \mathbf{e} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{e}, x_i \in \{0,1\} \\ & \underbrace{\mathbf{Y_S} \le \mathbf{XS}}_{y_{i,j} \le x_i}, \underbrace{\mathbf{Y_S} \le \mathbf{SX}}_{y_{i,j} \le x_j} \end{aligned}$$

- solve relaxed version
- more variables
- “big” LP solver required
- **upper bound** on solution

$\mathbf{x}$ still fractional, use $\alpha \mathbf{w} + \beta/2 \mathbf{Y_S} \mathbf{x}$ as the “weight” for MAX-WEIGHT MATCHING rounding.

NETALIGNLP	
Overlap	14,253
Weight	46,327 (96,297)

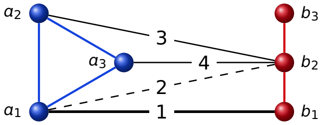
Upper bound on overlap 28660

Note *Took 1 day to compute*

Klau, BMC Bioinformatics 10:S59, 2009.

TIGHTLP

Trivial upper bound on overlap $\mathbf{e}^T \mathbf{S}\mathbf{e}/2 = 892,655$



$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$\mathbf{x}$ must be a matching eventually.

$$\mathbf{x}^T \mathbf{S}\mathbf{x} = \sum x_i \mathbf{s}_i^T \mathbf{x} \le \sum \mathbf{s}_i^T \mathbf{x} \le \sum \texttt{maxmatch}(\mathbf{s}_i, \mathbf{A}) = 344234.5$$

NETALIGN TIGHTLP	
Overlap	17,251
Weight	46,270 (96,141)

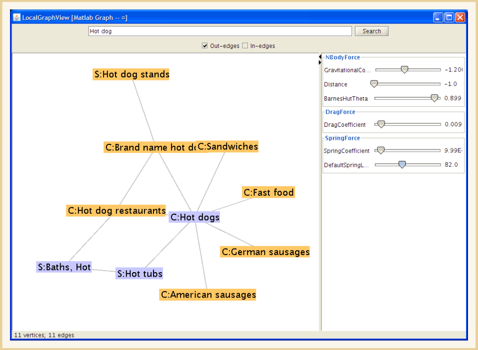
Upper bound on overlap 17,571

Note *Took 1 day to compute*

Klau, BMC Bioinformatics 10:S59, 2009.

Matching results

LCSH	WC
Science fiction television series	Science fiction television programs
Turing test	Turing test
Maching learning	Machine learning
Hot tubs	Hot dog



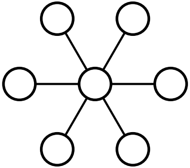
Graph similarity

Slide 21 of 29

DAVID’S GENERAL WARNING: SPECULATIVE AND UNVERIFIED. SOME MATERIAL MAY CONFUSE ME

Graph similarity

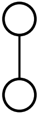
$\mathcal{A}$
big



$\mathcal{L}$
complete

"all edges"

$\mathcal{B}$
small



Make it a many to 1 matching

Graph similarity

Problem

$$\underset{X_{r,u}}{\text{maximize}} \quad \sum_{r,u} X_{r,u} W_{r,u} + \sum_{(r,u),(s,v)\in\mathcal{S}} X_{r,u} X_{s,v}$$

$$\text{subject to} \quad \underbrace{\sum_{u} X_{r,u} \le 1 \quad \forall r}_{\text{matching for } \mathcal{A}\to\mathcal{B}}$$

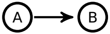
$$\longleftrightarrow \quad \begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \alpha \mathbf{w}^T \mathbf{x} + \frac{\beta}{2} \mathbf{x}^T \mathbf{S} \mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{e} \end{array}$$

Integer? Rounding?

Note For integer solutions, Yinyu Ye has a rounding procedure that would give you an approximation algorithm for the problem.

HITS as graph similarity

Idea

Given a graph $\mathcal{G}$, look at the similarity of $\mathcal{G}$ to 

Formulation

Largest eigenvalue of $\mathbf{S} = \mathbf{A}^T \otimes \mathbf{B}^T + \mathbf{A} \otimes \mathbf{B}$ gives HITS scores,

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \mathbf{x}^T \mathbf{S} \mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|_2 = 1 \end{array}$$

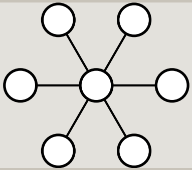
Weaknesses

- if **A** or **B** is symmetric, then **X** has rank 1
- no way to handle “sparse” case

We can do the same thing too!

Should I have cited Kleinberg too? : -) Blondel et al. SIREV 2004.

Examples

$\mathcal{A}$	$\mathcal{B}$	$\mathbf{X}$
		$\begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$
		$\begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$
`(rand(10)>0.5)`	`(rand(10)>0.5)`	$0.1\mathbf{e}\mathbf{e}^T$

Note *Studied already? Close to* PATTERN RECOGNITION *formulation.*
When $\mathcal{B} =$  *seems like* MAXCUT*, or finding a bipartition.*

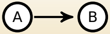
Sports ranking

Network alignment works on directed graphs too.

(Just define squares carefully.)

1. Align team networks against the dominance graph

$$\begin{pmatrix} 0 & 1 & \dots & 1 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

2. Evaluate similarity against the dominance graph (allows ties)
3. Evaluate similarity against 

Example

	Dav	UT	Cit	Cof	Wof	Sam	App	Elo	W.	Fur	UNC	GA
Dav		14	15	4	20	12	17	14	24	28	21	32
UT			6	11	7	9	4	12	7	27	14	13
Cit	18			12		25	4	2	12	13	7	15
Cof	5	9			2	7	6	8	15	9	15	13
Wof		3	5			11	6	3	15	14	12	7
Sam			9				11	13	12	8	14	23
App		6			5			19	6	4	1	12
Elo		1	2	6	7				6	8		9
W.		9		2	12	7	5	7		10	9	14
Fur		2				6					5	10
UNC							2	13				8
GA		4			21	7				16	9	

NETALIGN TIGHTLP

Overlap 64 Upper/Lower 739/197

Upper bound on overlap 65

Note Thanks to Kathryn Pendings for sending the data on short notice!

Summary

Network alignment

- Motivation from LCSH to Wikipedia
- QP, LP, and TightLP formulations
- Results (and problems) with LCSH to Wikipedia

Graph similarity

- "Network alignment" inspired formulation
- A few small examples

Sports ranking

- Interesting ideas for ranking teams?

When you have a hammer, everything looks like a nail.