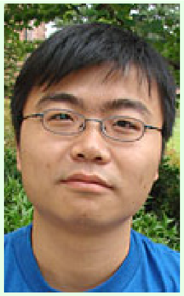
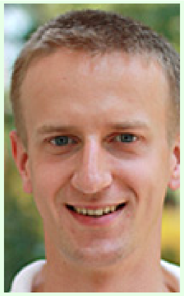
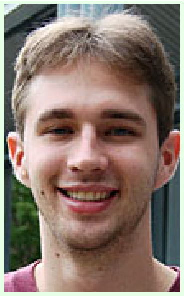
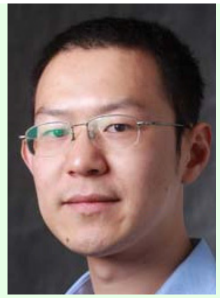

Mingyu Guo

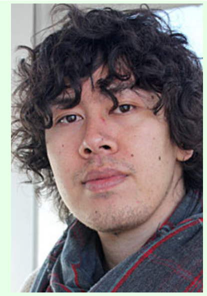

Dima Korzhyk


Josh Letchford


Liad Wagman


Lirong Xia


Troels Sorensen


Taiki Todo


Joe Farfel

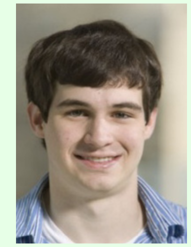

Melissa Dalis


Peng Shi

COMPUTATIONAL SOCIAL CHOICE

A Journey from Basic Complexity Results to a Brave New World for Social Choice

Vincent Conitzer, Duke University


Matt Rognlie

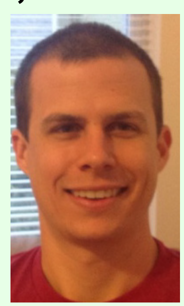

Bo Waggoner


Garrett Andersen

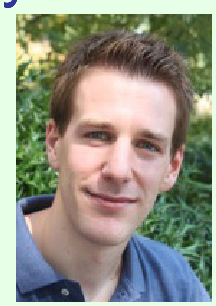

Rupert Freeman


Andrew Kephart


Yuqian Li

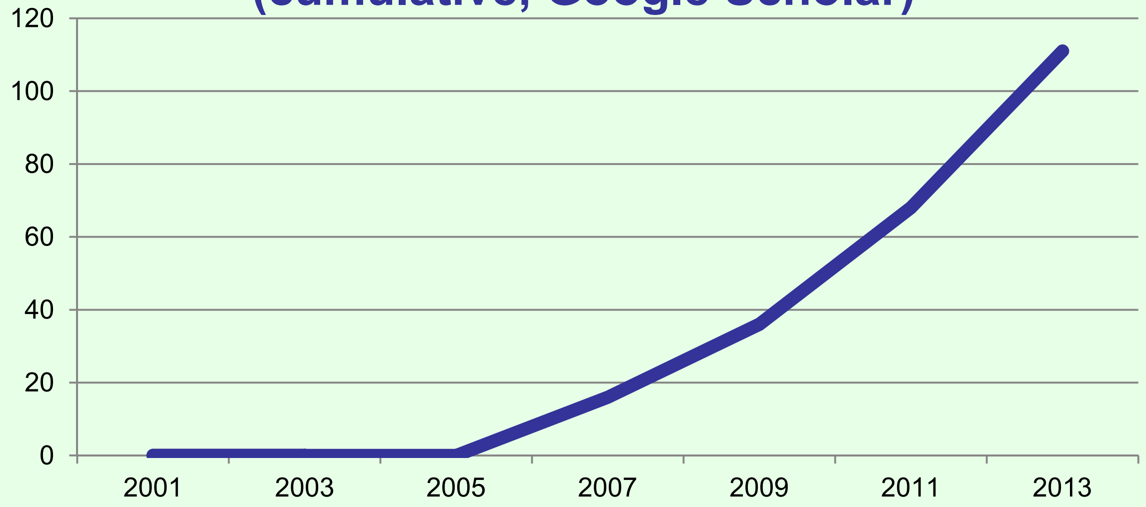

Aaron Kolb


Catherine Moon


Markus Brill


Angelina Vidali

A brief history of computational social choice



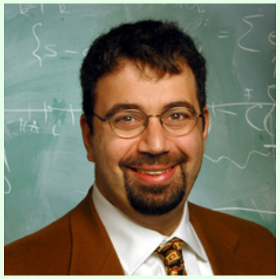
- Two 1989 papers by John Bartholdi, III, Craig Tovey, and Michael Trick
  - Voting schemes for which it can be difficult to tell who won the election. Social Choice and Welfare, 6:157-165.
  - The computational difficulty of manipulating an election. Social Choice and Welfare, 6:227-241.

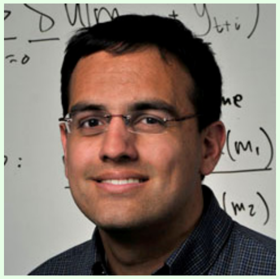
me in ~1989

(thanks mom)

Voting

n voters…





… each produce a ranking of *m* alternatives…

$b \succ a \succ c$

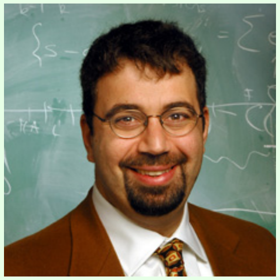
$a \succ c \succ b$

$a \succ b \succ c$

… which a social preference function maps to one or more aggregate rankings.

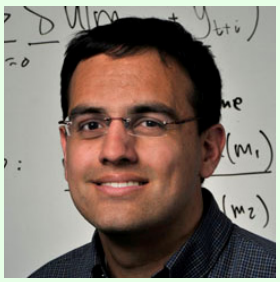
$a \succ b \succ c$

Kemeny



$b \succ a \succ c$

$a \succ b \succ c$



$a \succ c \succ b$

2 disagreements

$\leftrightarrow$

3*3 - 2 = 7 agreements (maximum)

$a \succ b \succ c$

- The unique SPF satisfying neutrality, consistency, and the Condorcet property [Young & Levenglick 1978]
- Natural interpretation as maximum likelihood estimate of the "correct" ranking [Young 1988, 1995]

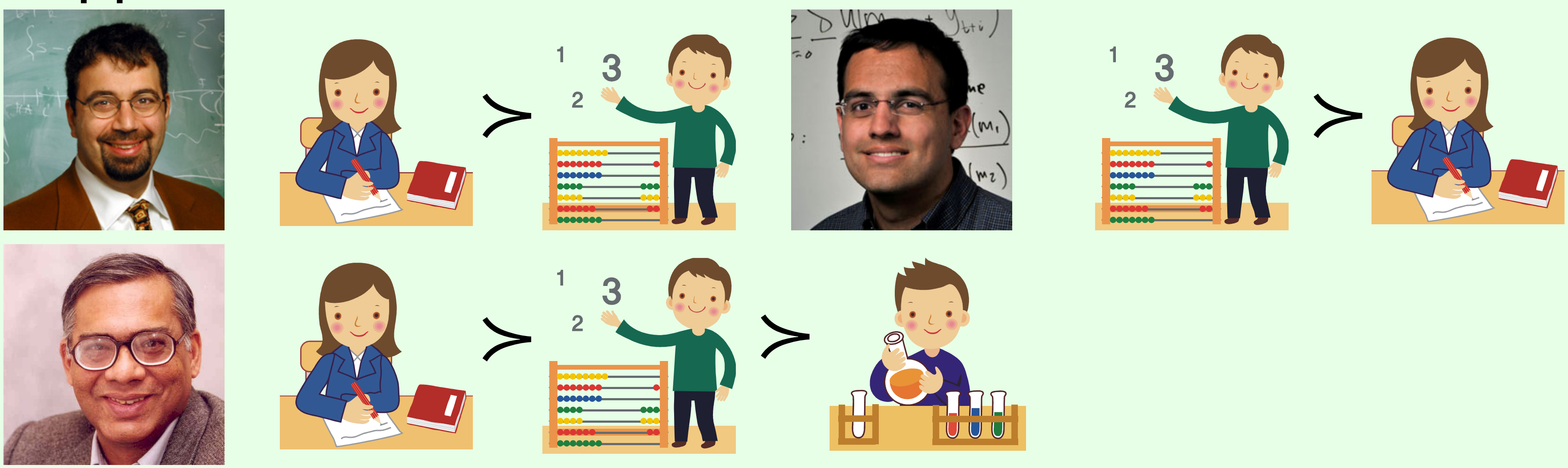
Objectives of voting

- **OBJ$_1$:** Compromise among subjective preferences

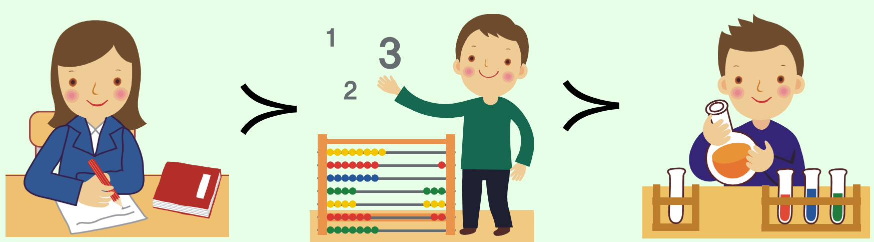
- **OBJ$_2$:** Reveal the "truth"

Ranking Ph.D. applicants
(briefly described in C. [2010])

- Input: Rankings of subsets of the (non-eliminated) applicants

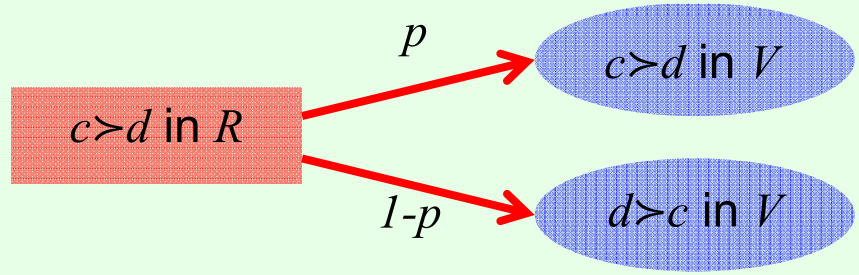


- Output: (one) Kemeny ranking of the (non-eliminated) applicants



An MLE model [dating back to Condorcet 1785]

- Correct outcome is a ranking R , $p>1/2$

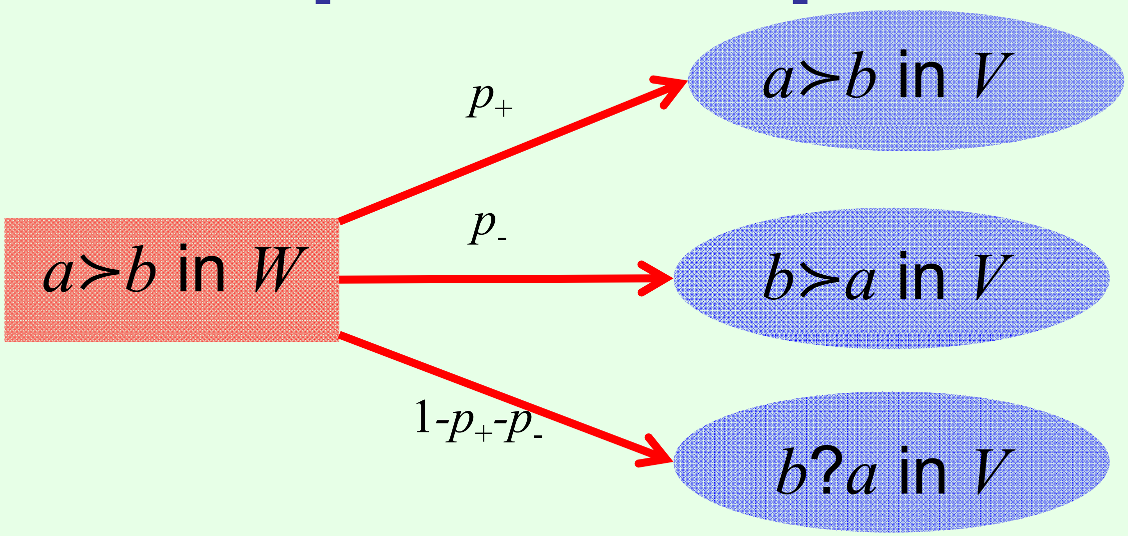


$$\Pr(\ b \succ c \succ a \mid a \succ b \succ c\ ) = \ ?\ p(1-p)^2$$

- MLE = Kemeny rule [Young 1988, 1995]
- Various other rules can be justified with different noise models [Drissi-Bakhkhat & Truchon 2004, C. & Sandholm 2005, Truchon 2008, C., Rognlie, Xia 2009, Procaccia, Reddi, Shah 2012]
  - 15:30 today: MLE in voting on social networks

A variant for partial orders

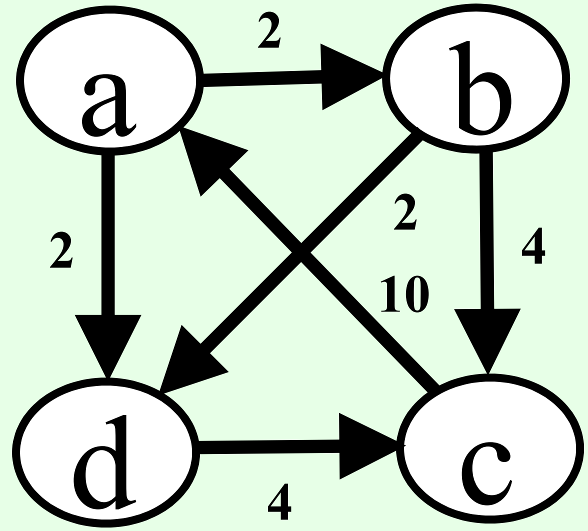
[Xia & C. 2011]



- Still gives Kemeny as the MLE

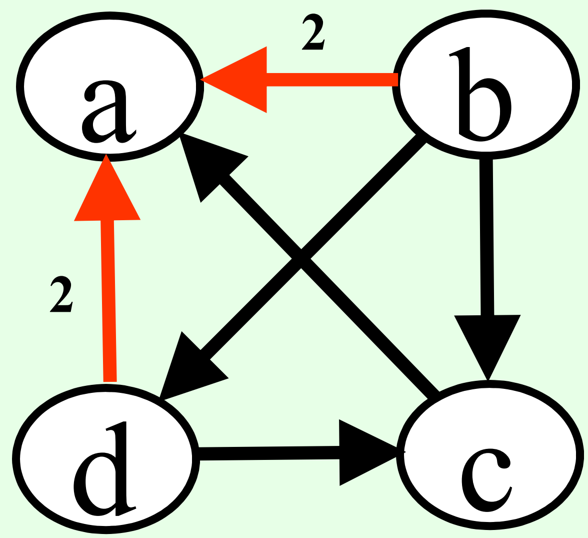
Computing Kemeny rankings

- 2 times $a \succ b \succ d \succ c$
- 5 times $a \succ d \succ b \succ c$
- 7 times $b \succ d \succ c \succ a$
- 6 times $c \succ a \succ d \succ b$
- 4 times $c \succ b \succ d \succ a$



- Final ranking = acyclic tournament graph
  - Edge (a, b) means a ranked above b
  - Acyclic = no cycles, tournament = edge between every pair
- Kemeny ranking seeks to minimize the total weight of the inverted edges
  - (minimizing their number = Slater)

Kemeny ranking



($b \succ d \succ c \succ a$)

A simple integer program for computing Kemeny rankings

(see, e.g., C., Davenport, Kalagnanam [2006])

Variable $x_{(a, b)}$ is 1 if *a* is ranked above *b*, 0 otherwise

Parameter $w_{(a, b)}$ is the weight on edge *(a, b)*

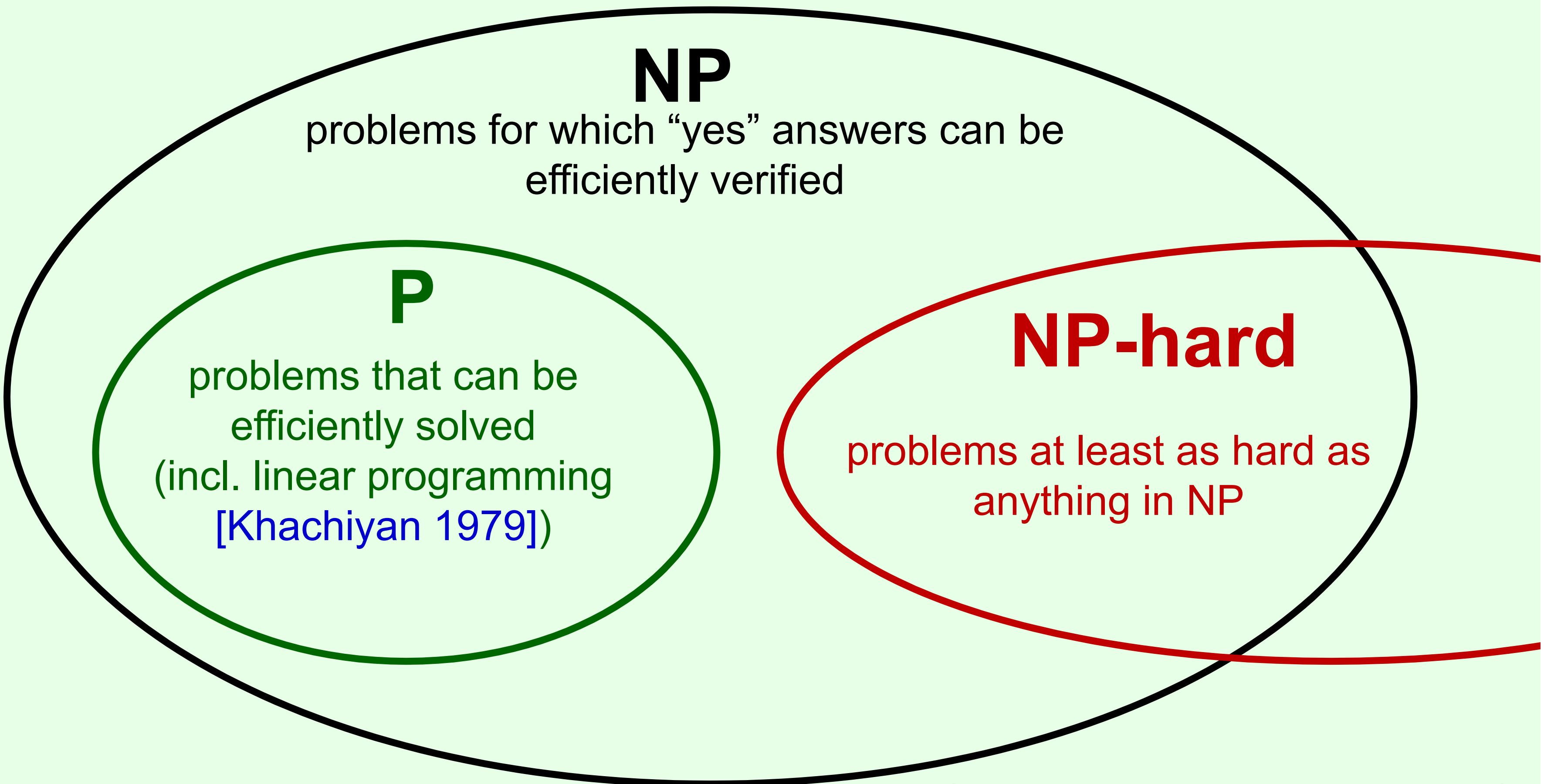
maximize: $\Sigma_{e \in E}\, w_e\, x_e$

subject to:

for all a, b $\in A$, $x_{(a, b)} + x_{(b, a)} = 1$

for all a, b, c $\in A$, $x_{(a, b)} + x_{(b, c)} + x_{(c, a)} \leq 2$

Computational complexity theory



(This picture assumes P ≠ NP.)

P = NP? [Cook 1971, Karp 1972, Levin 1973, …]

Complexity of Kemeny (and Slater)

- Kemeny:

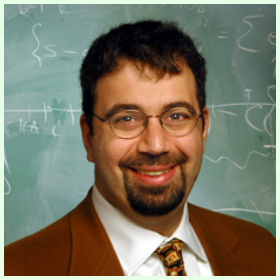
  NP-hard [Bartholdi, Tovey, Trick 1989]

  - Even with only 4 voters [Dwork, Kumar, Naor, Sivakumar 2001]
  - Exact complexity of Kemeny winner determination: complete for Θ_2^p [Hemaspaandra, Spakowski, Vogel 2005]

- Slater:

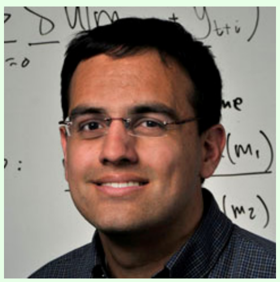
  - NP-hard, even if there are no pairwise ties [Ailon, Charikar, Newman 2005, Alon 2006, C. 2006, Charbit, Thomassé, Yeo 2007]

Instant runoff voting / single transferable vote (STV)



$b \succ a \succ c$

$a \succ b \succ c$



$a \succ b \succ b$

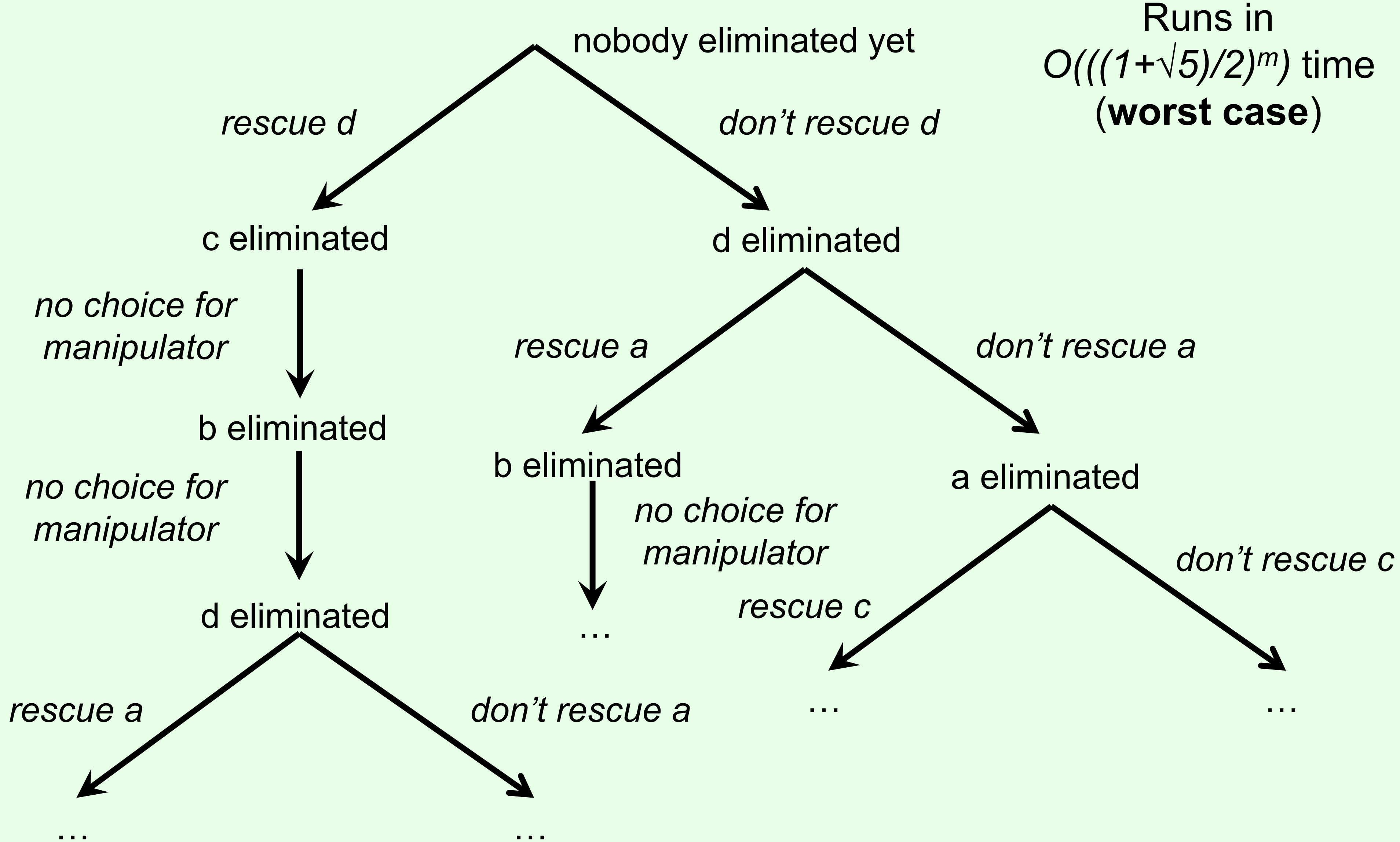
$a \succ b \succ c$

- The unique SPF satisfying: independence of bottom alternatives, consistency at the bottom, independence of clones (& some minor conditions) [Freeman, Brill, C. 2014 – 11am today]
- NP-hard to manipulate [Bartholdi & Orlin, 1991]

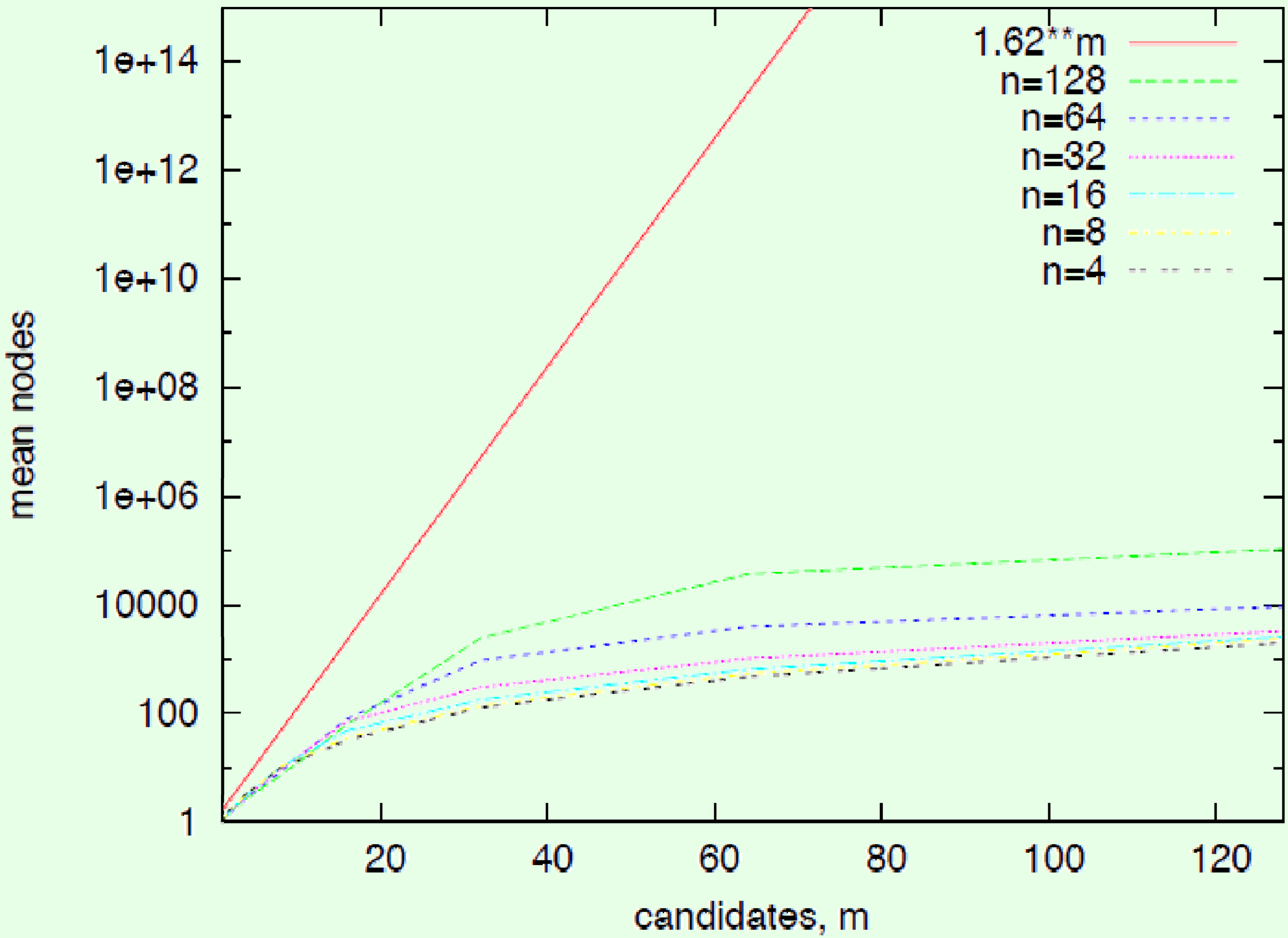
STV manipulation algorithm

[C., Sandholm, Lang 2007]

Runs in $O(((1+\sqrt{5})/2)^m)$ time (**worst case**)



Runtime on random votes [Walsh 2011]

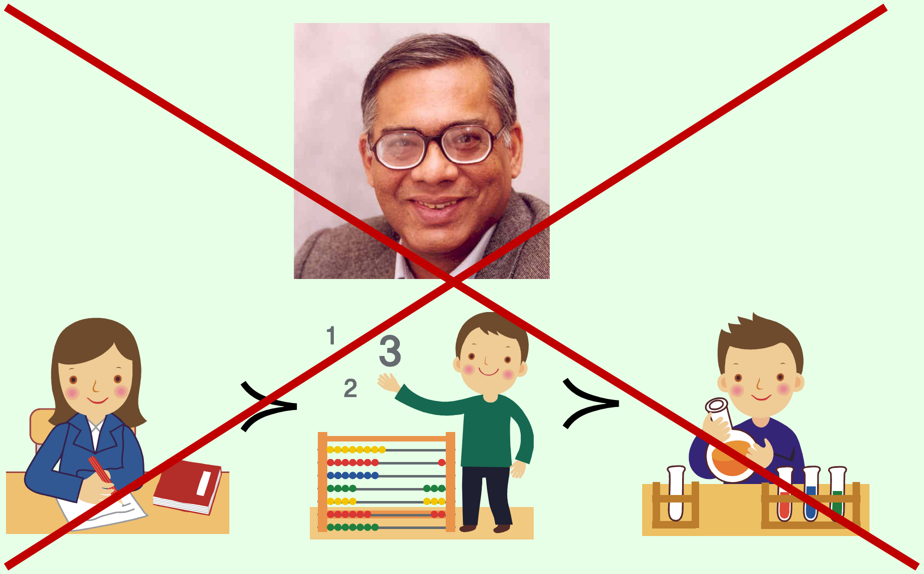


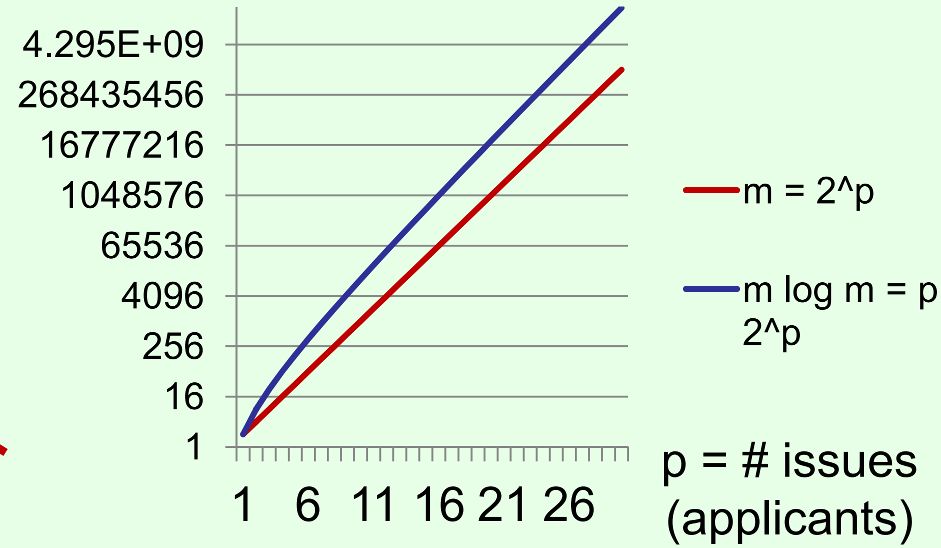
Fine – how about another rule?

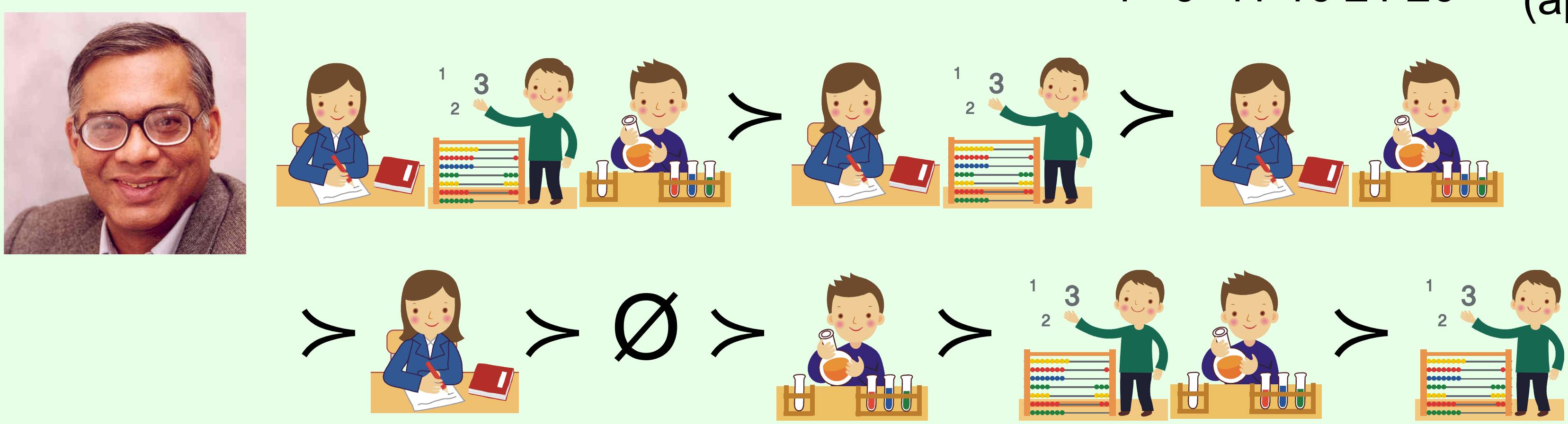
- Heuristic algorithms and/or experimental (simulation) evaluation [C. & Sandholm 2006, Procaccia & Rosenschein 2007, Walsh 2011, Davies, Katsirelos, Narodytska, Walsh 2011]
- Quantitative versions of Gibbard-Satterthwaite showing that under certain conditions, for some voter, even a random manipulation on a random instance has significant probability of succeeding [Friedgut, Kalai, Nisan 2008; Xia & C. 2008; Dobzinski & Procaccia 2008; Isaksson, Kindler, Mossel 2010; Mossel & Racz 2013]

"for a social choice function f on k≥3 alternatives and n voters, which is ϵ-far from the family of nonmanipulable functions, a uniformly chosen voter profile is manipulable with probability at least inverse polynomial in n, k, and ϵ^{-1}."

Ph.D. applicants may be substitutes or complements…







Sequential voting and strategic voting

S | **T**

$$V_1 : st > \bar{s}t > s\bar{t} > \bar{s}\bar{t}$$
$$V_2 : s\bar{t} > st > \bar{s}t > \bar{s}\bar{t}$$
$$V_3 : \bar{s}t > \bar{s}\bar{t} > s\bar{t} > st$$

- In the first stage, the voters vote simultaneously to determine **S**; then, in the second stage, the voters vote simultaneously to determine **T**
- If **S** is built, then in the second step $t > \bar{t}$, $\bar{t} > t$, $\bar{t} > t$ so the winner is $s\bar{t}$
- If **S** is **not** built, then in the 2nd step $t > \bar{t}$, $t > \bar{t}$, $t > \bar{t}$ so the winner is $\bar{s}t$
- In the first step, the voters are effectively comparing $s\bar{t}$ and $\bar{s}t$, so the votes are $\bar{s} > s$, $s > \bar{s}$, $\bar{s} > s$, and the final winner is $\bar{s}t$

[Xia, C., Lang 2011; see also Farquharson 1969, McKelvey & Niemi 1978, Moulin 1979, Gretlein 1983, Dutta & Sen 1993]

Multiple-election paradoxes for strategic voting [Xia, C., Lang 2011]

- **Theorem** (informally). For any $p \geq 2$ and any $n \geq 2p^2 + 1$, there exists a profile such that the strategic winner is
  - ranked almost at the bottom (exponentially low positions) in every vote
  - Pareto dominated by almost every other alternative
  - an almost Condorcet loser

- Multiple-election paradoxes [Brams, Kilgour & Zwicker 1998], [Scarsini 1998], [Lacy & Niou 2000], [Saari & Sieberg 2001], [Lang & Xia 2009], [C. & Xia 2012]

Time Magazine "Person of the Century" poll – "results" (January 19, 2000)

#	Person	%	Tally
1	Elvis Presley	13.73	625045
2	Yitzhak Rabin	13.17	599473
3	Adolf Hitler	11.36	516926
4	Billy Graham	10.35	471114
5	**Albert Einstein**	**9.78**	**445218**
6	Martin Luther King	8.40	382159
7	Pope John Paul II	8.18	372477
8	Gordon B Hinckley	5.62	256077
9	Mohandas Gandhi	3.61	164281
10	Ronald Reagan	1.78	81368
11	John Lennon	1.41	64295
12	American GI	1.35	61836
13	Henry Ford	1.22	55696
14	Mother Teresa	1.11	50770
15	Madonna	0.85	38696
16	Winston Churchill	0.83	37930
17	Linus Torvalds	0.53	24146
18	Nelson Mandela	0.47	21640
19	Princess Diana	0.36	16481
20	Pope Paul VI	0.34	15812

Time Magazine "Person of the Century" poll – partial results (November 20, 1999)

#	Person	%	Tally
1	Jesus Christ	48.36	610238
2	Adolf Hitler	14.00	176732
3	Ric Flair	8.33	105116
4	Prophet Mohammed	4.22	53310
5	John Flansburgh	3.80	47983
6	Mohandas Gandhi	3.30	41762
7	Mustafa K Ataturk	2.07	26172
8	Billy Graham	1.75	22109
9	Raven	1.51	19178
10	Pope John Paul II	1.15	14529
11	Ronald Reagan	0.98	12448
12	Sarah McLachlan	0.85	10774
13	Dr William L Pierce	0.73	9337
14	Ryan Aurori	0.60	7670
15	Winston Churchill	0.58	7341
16	**Albert Einstein**	**0.56**	**7103**
17	Kurt Cobain	0.32	4088
18	Bob Weaver	0.29	3783
19	Bill Gates	0.28	3629
20	Serdar Gokhan	0.28	3627

Anonymity-proof voting rules

- A voting rule is **false-name-proof** if no voter ever benefits from participating more than once
  - Studied in combinatorial auctions by Yokoo, Sakurai, Matsubara [2004] (inefficiency ratio by Iwasaki, C., Omori, Sakurai, Todo, Guo, Yokoo [2010]); in matching by Todo & C. [2013]
- A voting rule **satisfies voluntary participation** if it never hurts a voter to cast her vote
- A voting rule is **anonymity-proof** if it is false-name-proof & satisfies voluntary participation
- Can we characterize (neutral, anonymous, randomized) anonymity-proof rules?

Anonymity-proof voting rules - characterization

- **Theorem** [C. 2008] (cf. Gibbard [1977] for strategy-proof randomized rules) :

  Any anonymity-proof (neutral, anonymous) voting rule f can be described by a single number p_f in [0,1]

  With probability p_f, the rule chooses an alternative uniformly at random

  With probability 1- p_f, the rule draws two alternatives uniformly at random;

  – if all votes rank the same alternative higher among the two, that alternative is chosen

  – otherwise, a fair coin is flipped to decide between the two alternatives.

- Assuming single-peaked preferences does not help much [Todo, Iwasaki, Yokoo 2011]

How should we deal with these negative results?

- Assume creating additional identifiers comes at a cost [Wagman & C. 2008]
- Verify some of the identities [C. 2007]
- Try to make voting multiple times difficult, analyze carefully using statistical techniques [Waggoner, Xia, C., 2012]
- Use social network structure [C., Immorlica, Letchford, Munagala, Wagman, 2010]

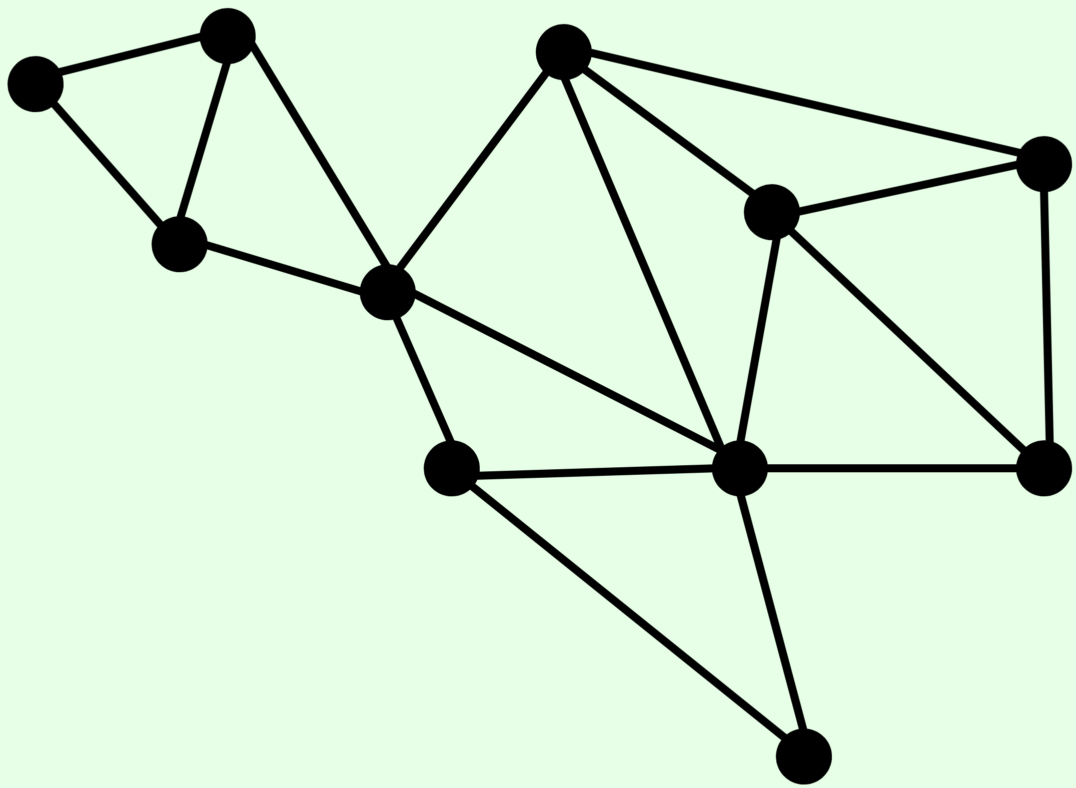
Facebook election

- In 2009, Facebook allowed its users to vote on its terms of use
  - Note: result would only be binding if >30% of its active users voted
  - #votes: ~600 000
  - #active users at the time: >200 000 000
- Could Facebook use its knowledge of the social network structure to prevent false-name manipulation?

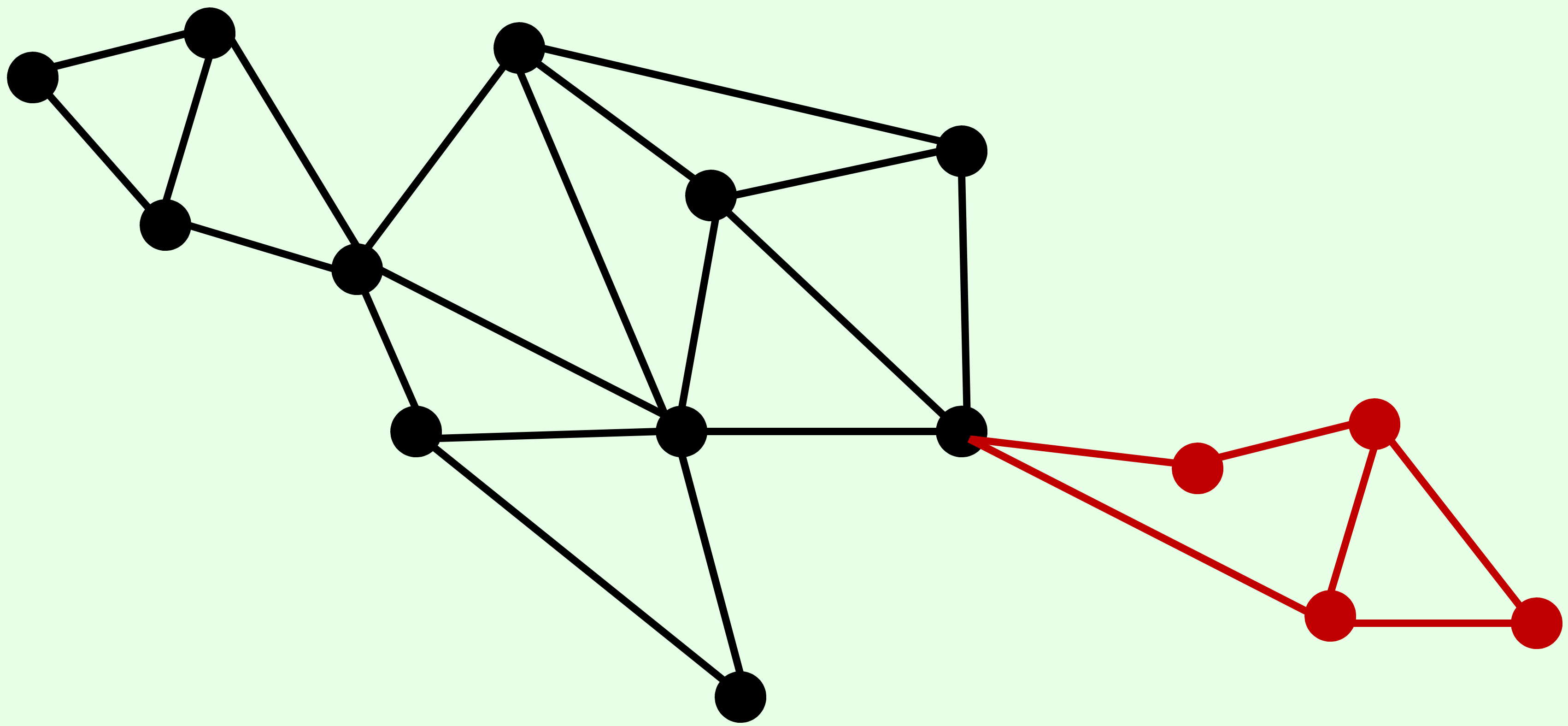
Related research

- Mostly in the systems community ("Sybil attacks") (e.g.: Yu, Gibbons, Kaminsky, Xiao [2010])
- Differences here:
  - rigorous mechanism design approach – should not benefit *at all* from creating false names
  - we allow things to be centralized

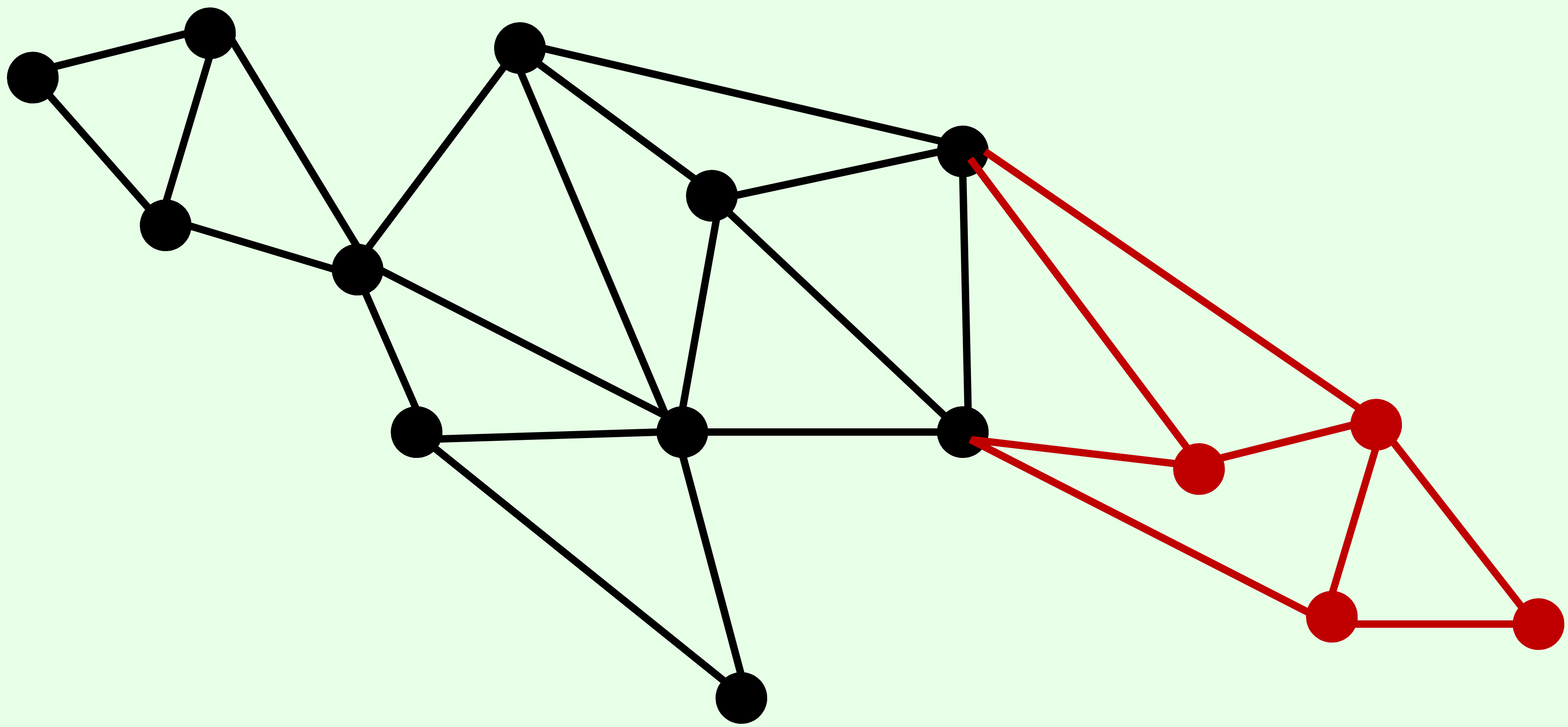
Social network graph



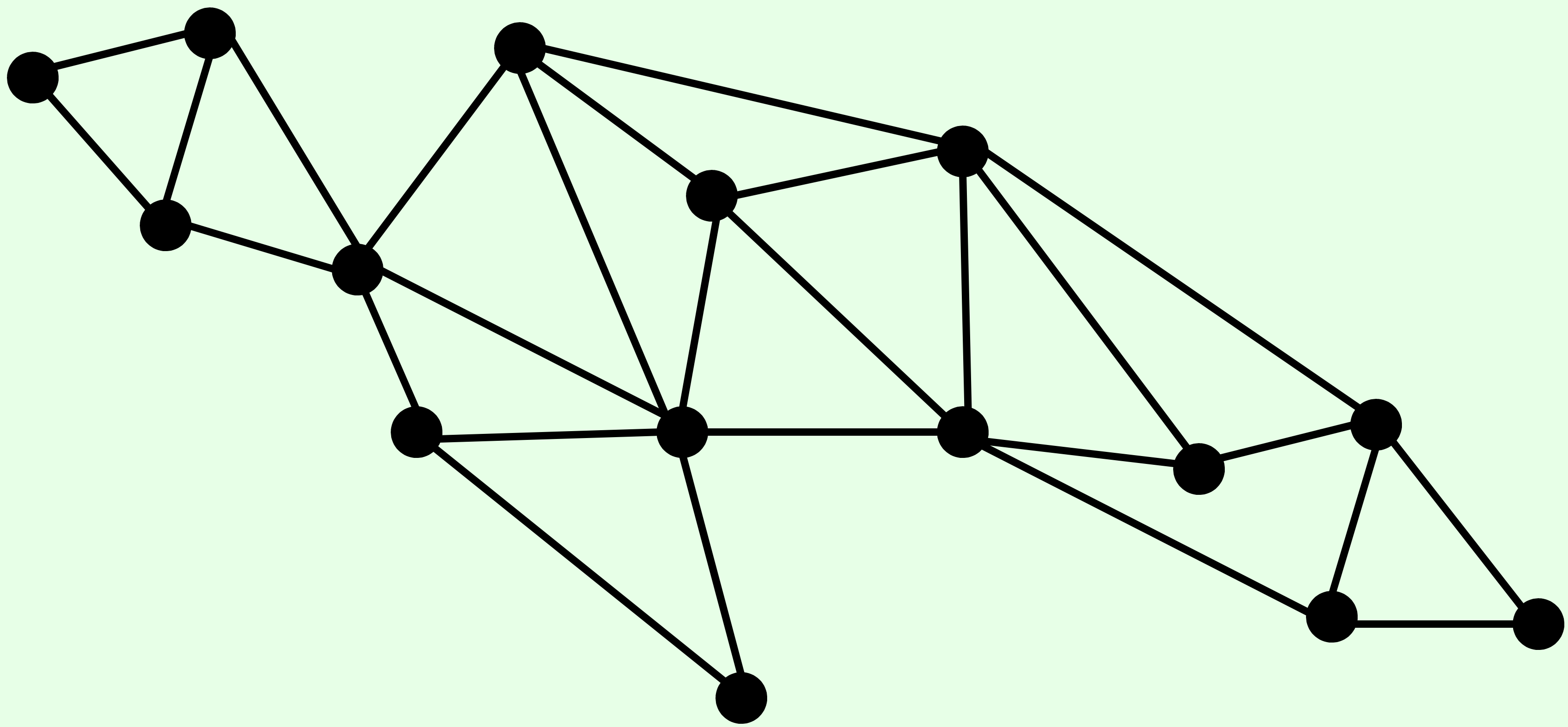
Creating new identities



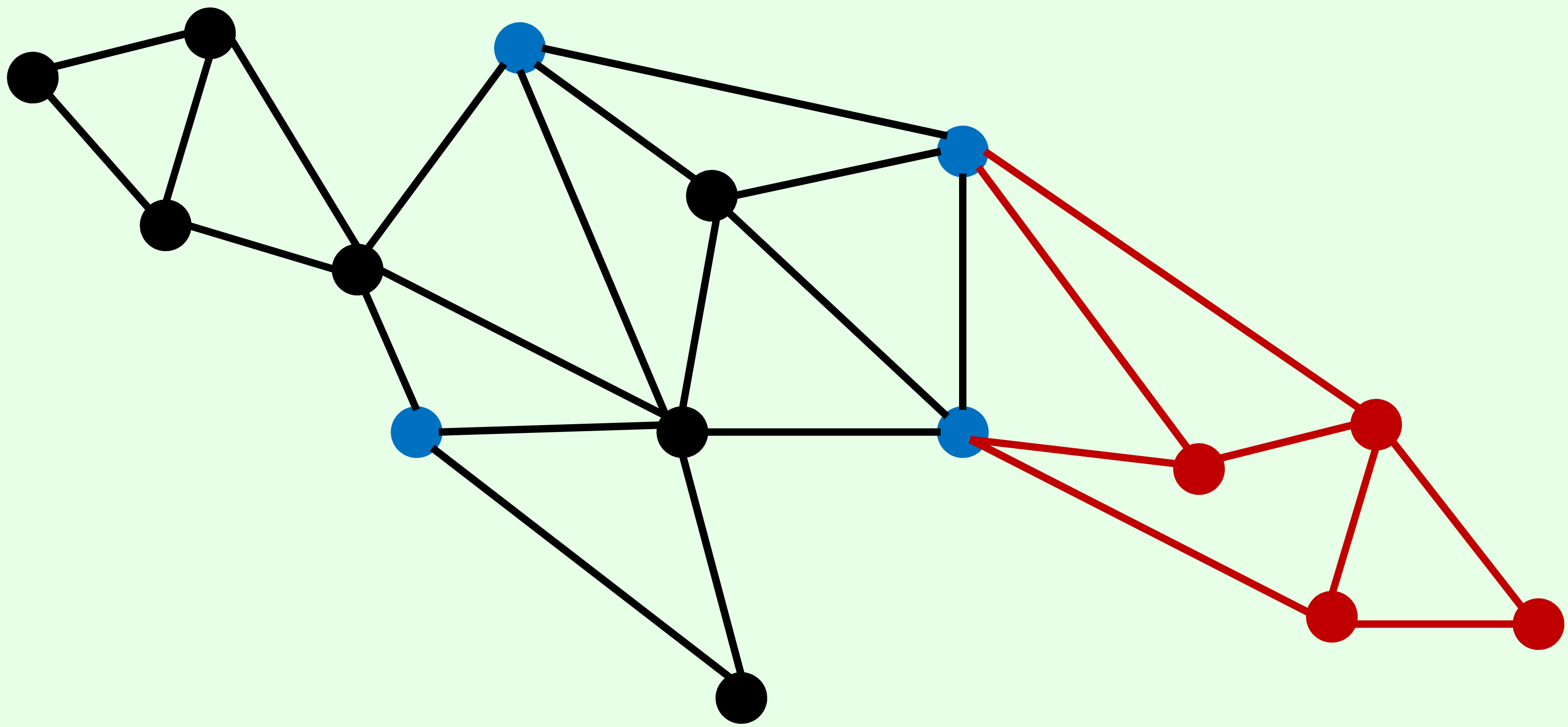
Coalitional manipulation



Election organizer's view

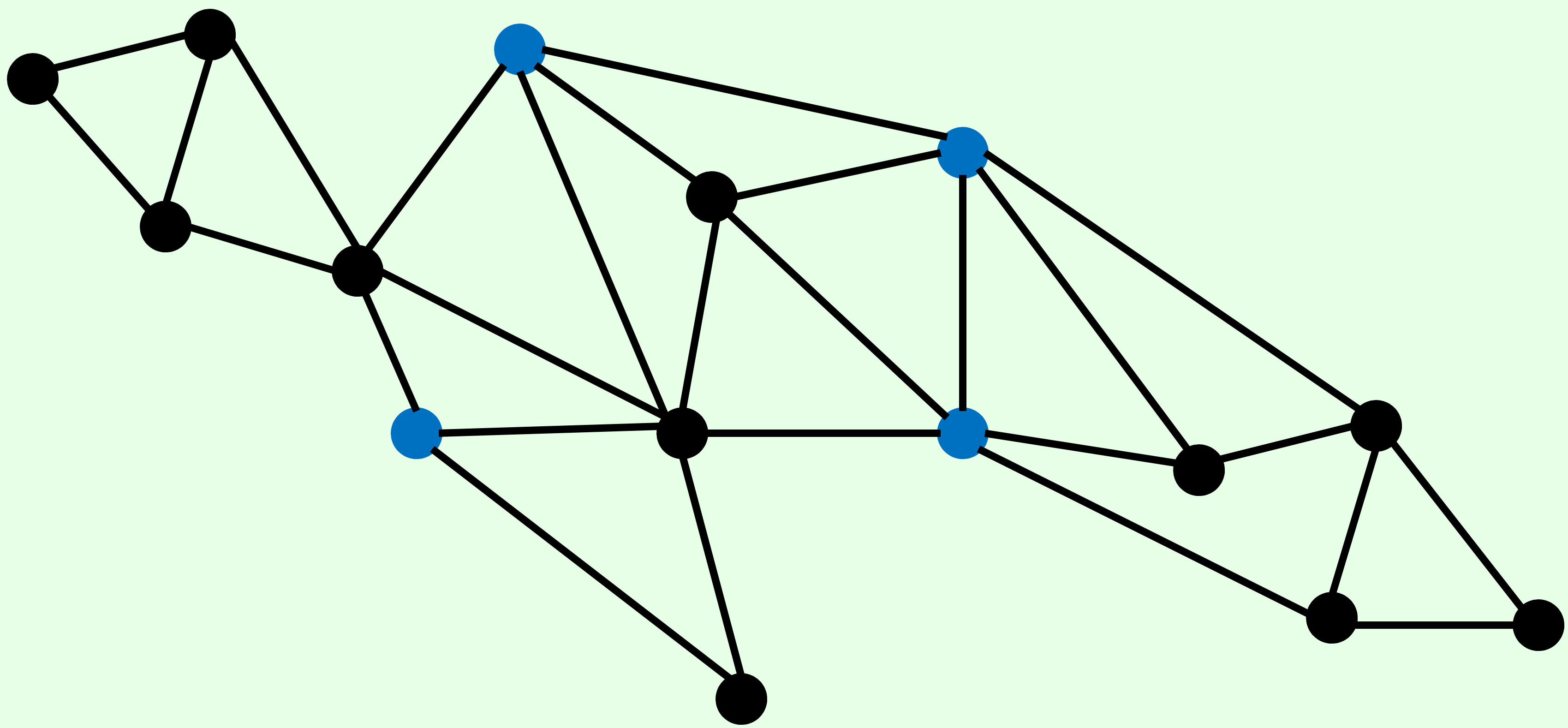


Trusted nodes



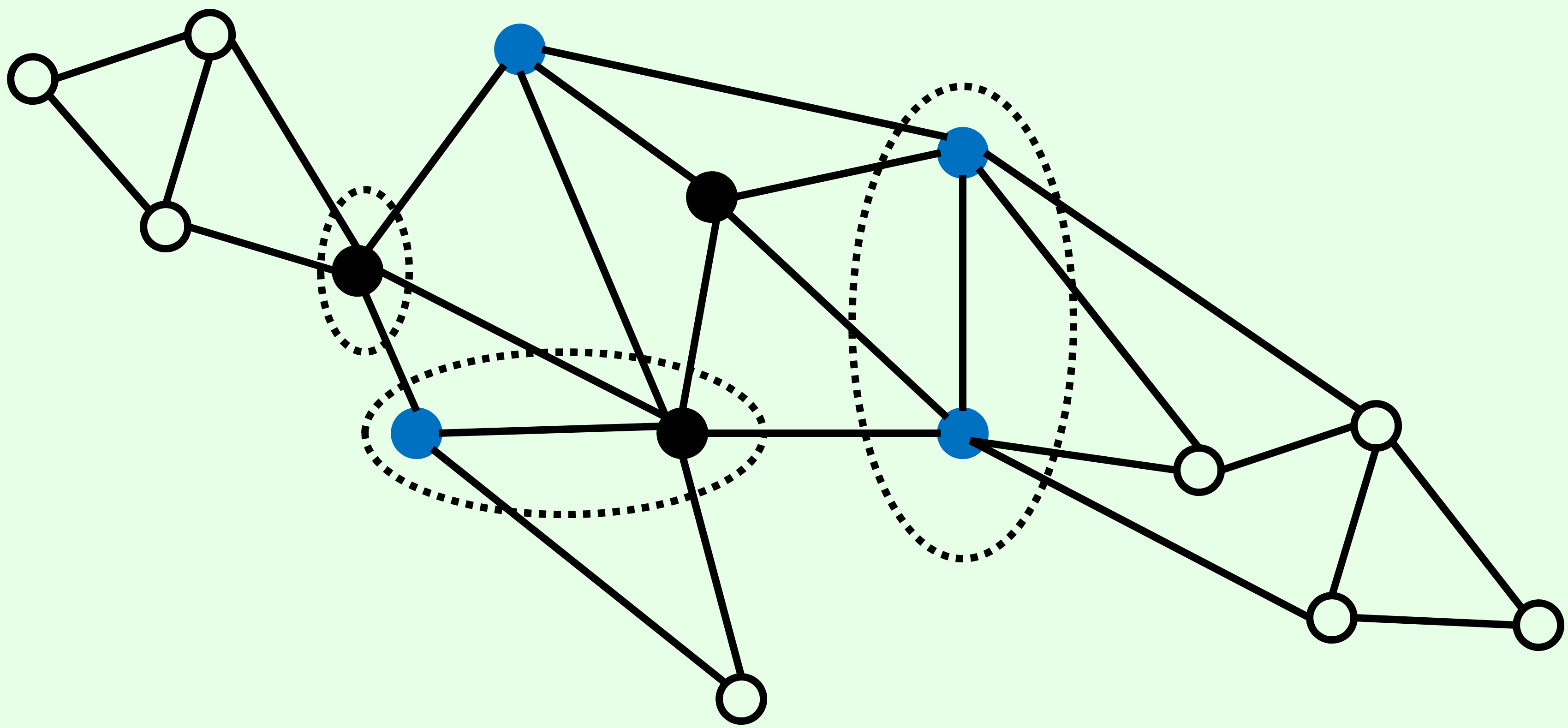
- Trusted nodes are known to be real, but may manipulate

Center's view



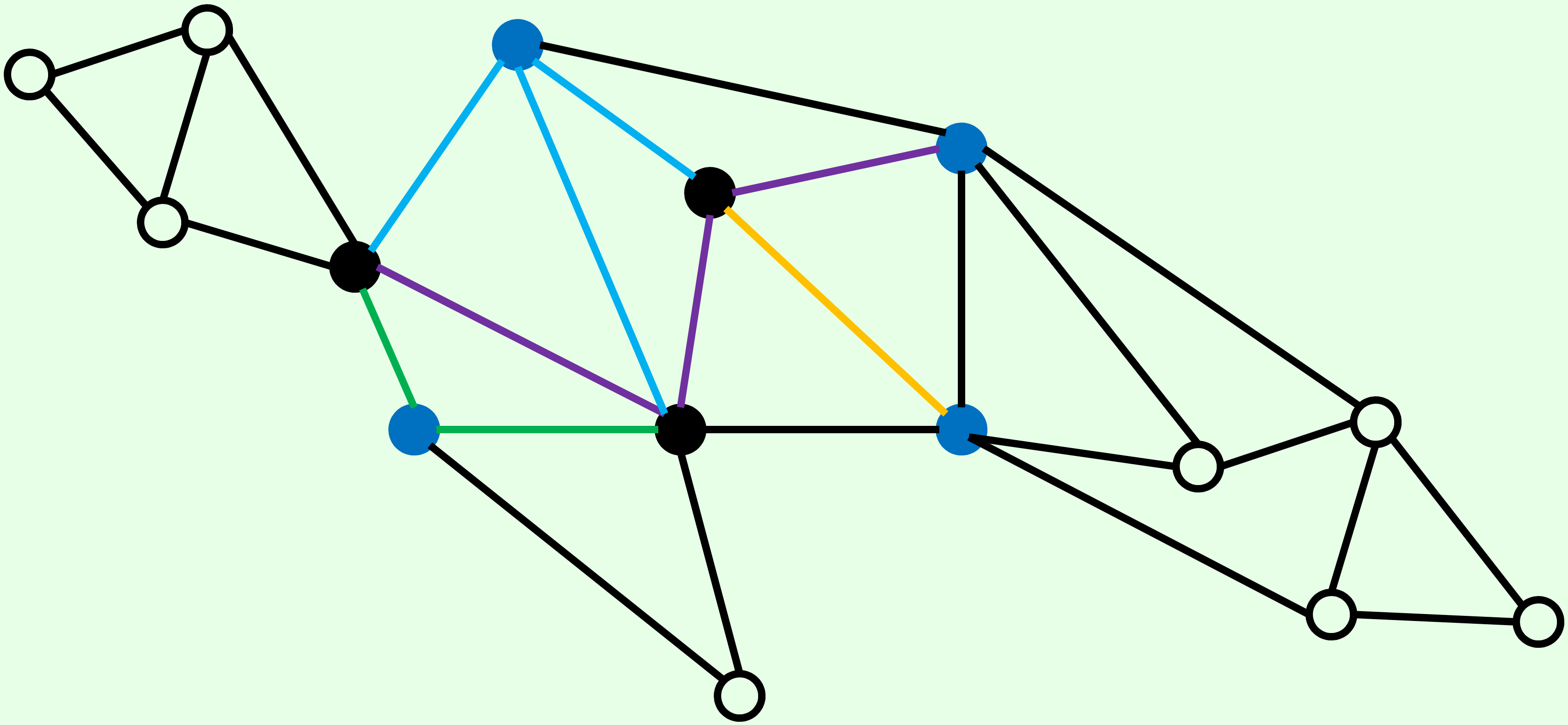
- Suppose the center knows that at most k legitimate nodes can work together (say, k=2)
- Which nodes can the center conclude are legitimate? Which are suspect?

Vertex cuts



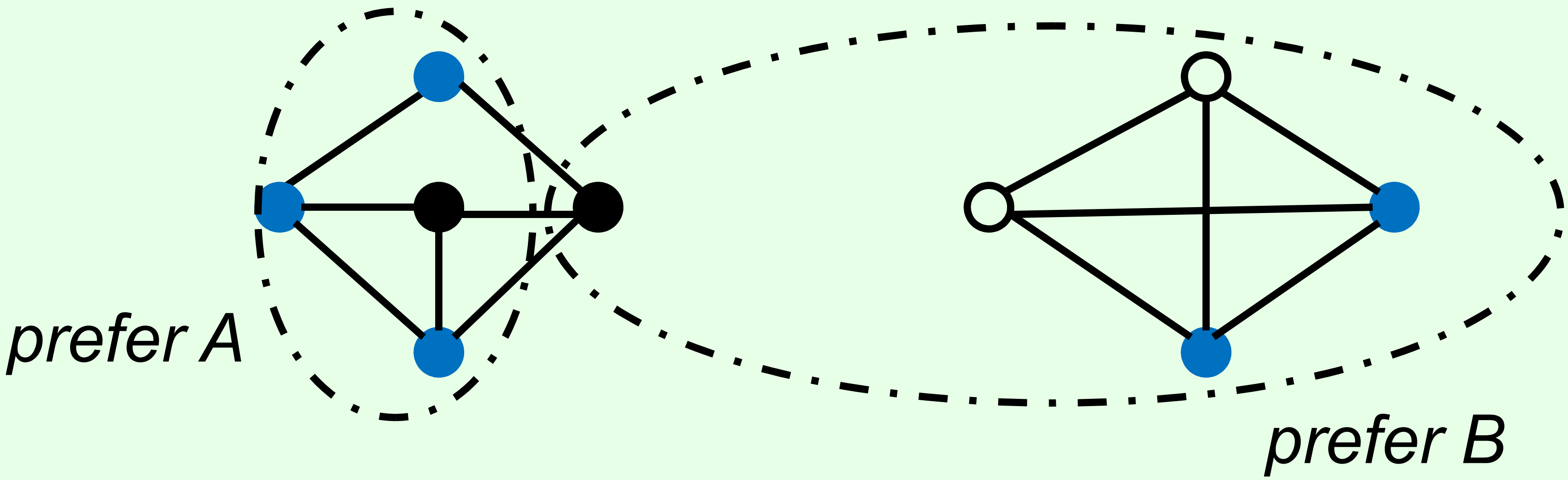
- Every node separated from the trusted nodes by a vertex cut of size at most k (=2) is suspect

Using Menger's theorem



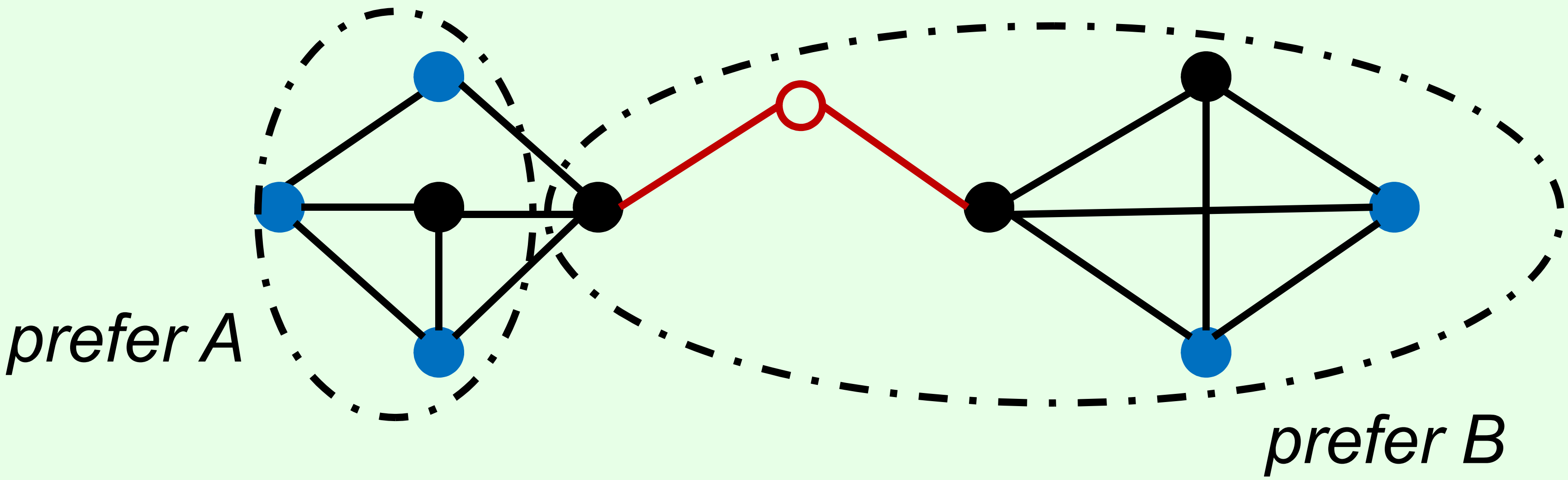
- A node *v* is not separated by a vertex cut of size at most *k* if and only if there are *k+1* vertex-disjoint paths from the trusted nodes to *v*
  - follows straightforwardly from Menger's theorem/duality

Is it enough to not let these suspect nodes vote? No…



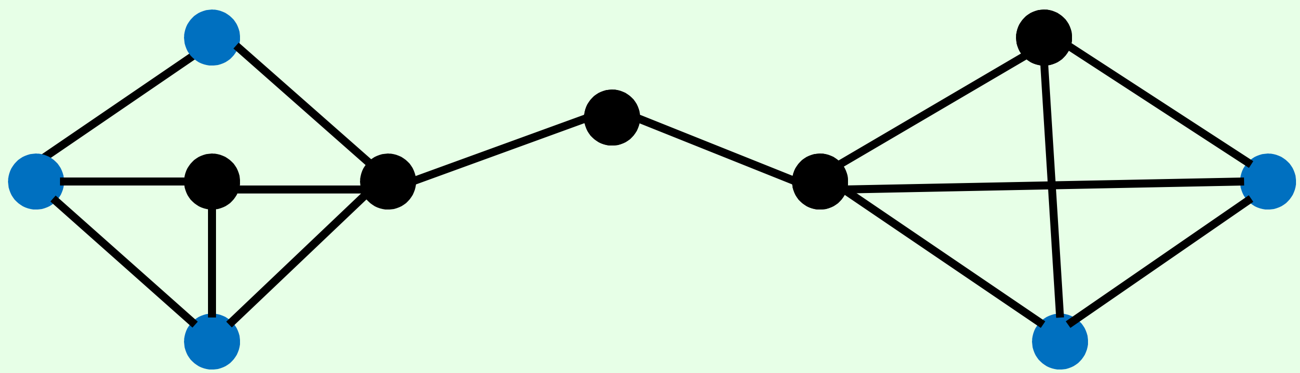
- Majority election between A and B, *k*=2
- A wins by 4 votes to 3 (two nodes don't get to vote for B)

Is it enough to not let these suspect nodes vote? No…



- Majority election between A and B, *k*=2
- B now wins by 5 votes to 4 (!)

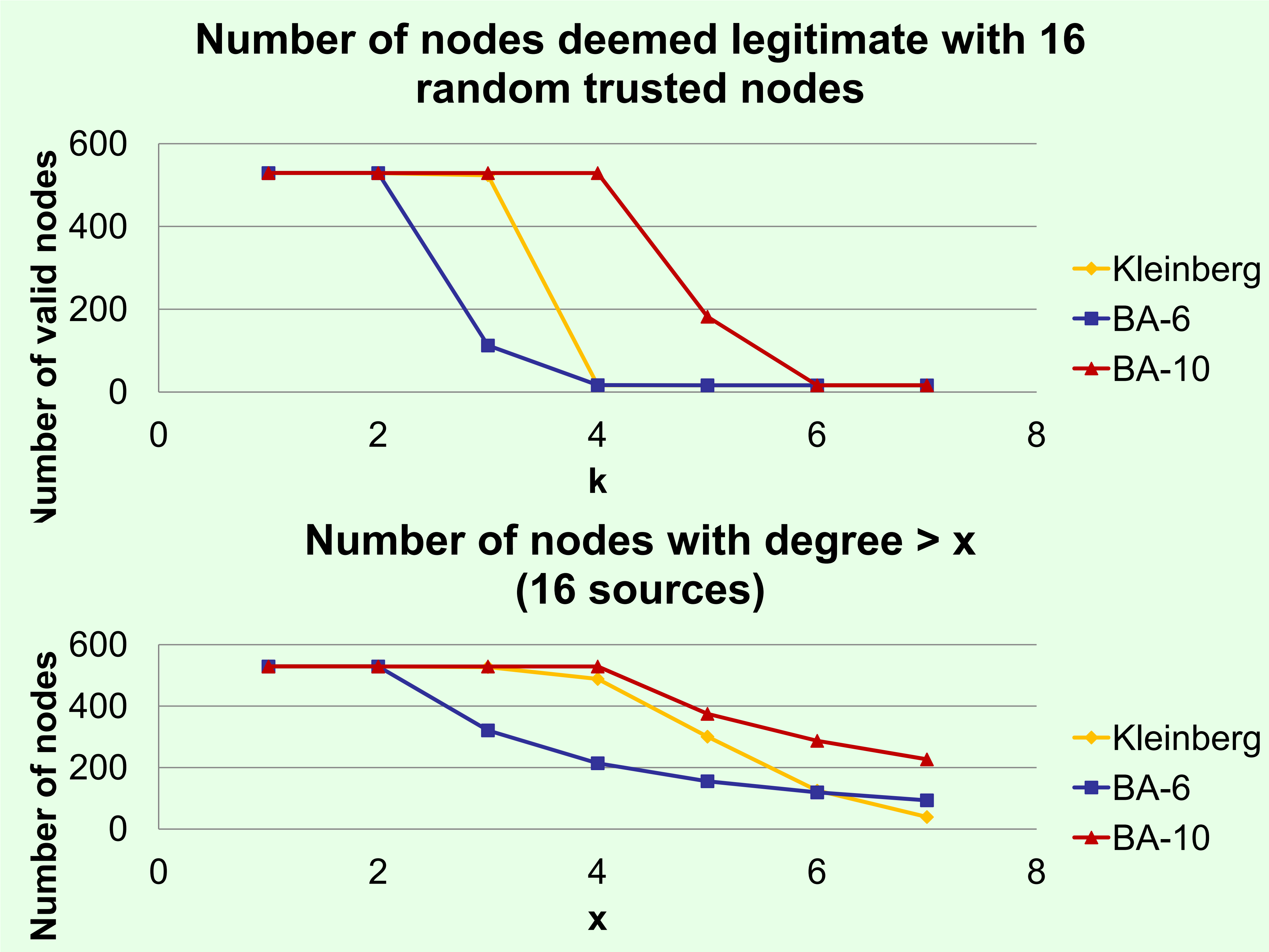
Solution: *iteratively* remove nodes separated by vertex cuts, until convergence



- Removes incentive for manipulation
- Call this suspicion policy *Π**

k-robustness

- Definition. A suspicion policy is *k*-robust if
  - the actions of one coalition of size at most *k* do not affect which nodes of other (disjoint) coalitions are deemed legitimate;
  - a coalition maximizes its number of identifiers that are deemed legitimate by not creating any false nodes.
- Theorem. A *k*-robust suspicion policy, combined with a standard mechanism that is both *k*-strategy-proof and satisfies *k*-voluntary participation, is false-name-proof for coalitions of size up to *k*.
- Theorem. Π^* is *k*-robust. Also, Π^* is guaranteed to label every illegitimate node as suspect. Finally, a coalition's false names do not affect which of its own legitimate nodes are deemed legitimate.
- Theorem. Any suspicion policy with these properties must label as suspect at least the nodes labeled as suspect by Π^*.



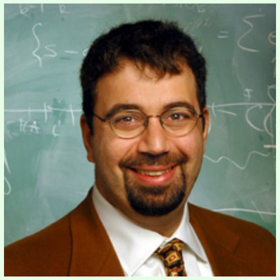
Number of nodes deemed legitimate with 16 random trusted nodes
Number of valid nodes
600
400
200
0
0
2
4
6
8
k
Kleinberg
BA-6
BA-10
Number of nodes with degree > x (16 sources)
Number of nodes
600
400
200
0
0
2
4
6
8
x
Kleinberg
BA-6
BA-10

Some shameless plugs:

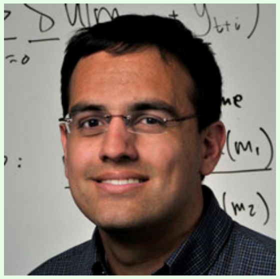
- COMSOC workshop starts this Monday in Pittsburgh!
- Computational social choice…
  - … mailing list: https://lists.duke.edu/sympa/subscribe/comsoc
  - … book: in preparation (editors: Brandt, C., Endriss, Lang, Procaccia)
  - … intro article: Brandt, C., Endriss [2013]
- New journal: *ACM Transactions on Economics and Computation (ACM TEAC)* (edited with Preston McAfee)

Thank you for your attention!

Bucklin



$b \succ a \succ c$



$a \succ c \succ b$

$a \succ b \succ c$

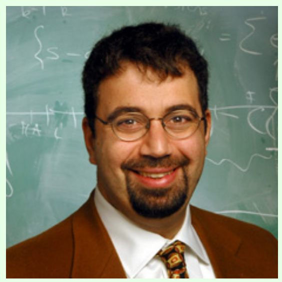
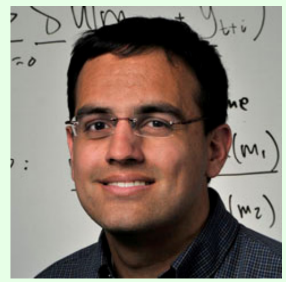
a's median rank: 1
b's median rank: 2
c's median rank: 3

$a \succ b \succ c$

An elicitation algorithm for the Bucklin voting rule based on binary search

[C. & Sandholm 2005]

- Alternatives: A B C D E F G H

- Top 4? {A B C D} {A B F G} {A C E H}
- Top 2? {A D} {B F} {C H}
- Top 3? {A C D} {B F G} {C E H}

Total communication is $nm + nm/2 + nm/4 + \ldots \leq 2nm$ bits
(n number of voters, m number of candidates)

Communication complexity

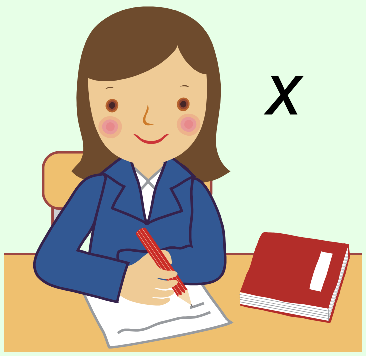
- Can also prove lower bounds on communication required for voting rules [C. & Sandholm 2005]

Rule	Lower bound	Upper bound
plurality	$\Omega(n \log m)$	$O(n \log m)$
plurality w/ runoff	$\Omega(n \log m)$	$O(n \log m)$
STV	$\Omega(n \log m)$	$O(n(\log m)^2)$
Condorcet	$\Omega(nm)$	$O(nm)$
approval	$\Omega(nm)$	$O(nm)$
Bucklin	$\Omega(nm)$	$O(nm)$
cup	$\Omega(nm)$	$O(nm)$
maximin	$\Omega(nm)$	$O(nm)$
Borda	$\Omega(nm \log m)$	$O(nm \log m)$
Copeland	$\Omega(nm \log m)$	$O(nm \log m)$
ranked pairs	$\Omega(nm \log m)$	$O(nm \log m)$

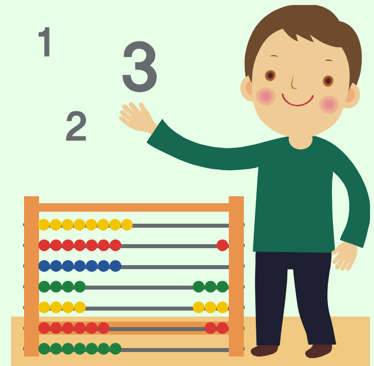
- Restrictions such as single-peaked preferences can help [C. 2009, Farfel & C. 2011]
- C. & Sandholm [2002]: strategic aspects of elicitation
- Service & Adams [2012]: communication complexity of *approximating* voting rules

Conditional preference networks (CP-nets)

[Boutilier, Brafman, Domshlak, Hoos, and Poole 2004]



x

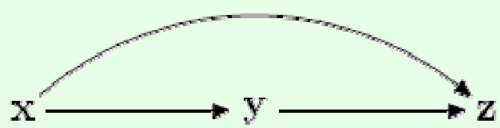


y

z

Variables: x, y, z. $D_x = \{x, \bar{x}\}, D_y = \{y, \bar{y}\}, D_z = \{z, \bar{z}\}$.

Directed graph, CPTs:



CPT(x)	CPT(y)	CPT(z)
$x \succ \bar{x}$	$x : y \succ \bar{y}$	$xy : z \succ \bar{z}$
	$\bar{x} : \bar{y} \succ y$	$x\bar{y} : z \succ \bar{z}$
		$\bar{x}y : z \succ \bar{z}$
		$\bar{x}\bar{y} : \bar{z} \succ z$

This CP-net encodes the following partial order:

$$xyz \nearrow x\bar{y}z \searrow,\quad xyz \searrow xy\bar{z} \nearrow \quad x\bar{y}\bar{z} \to \bar{x}\bar{y}\bar{z} \to \bar{x}\bar{y}z \to \bar{x}yz \to \bar{x}y\bar{z}$$

Sequential voting

see Lang & Xia [2009]

- Issues: main dish, wine
- Order: main dish > wine
- Local rules are majority rules
- V_1:  $\succ$  ,  :  $\succ$  ,  :  $\succ$ 
- V_2:  $\succ$  ,  :  $\succ$  ,  :  $\succ$ 
- V_3:  $\succ$  ,  :  $\succ$  ,  :  $\succ$ 
- **Step 1:** 
- **Step 2: given**  **,**  **is the winner for wine**
- **Winner: (**  **,**  **)**

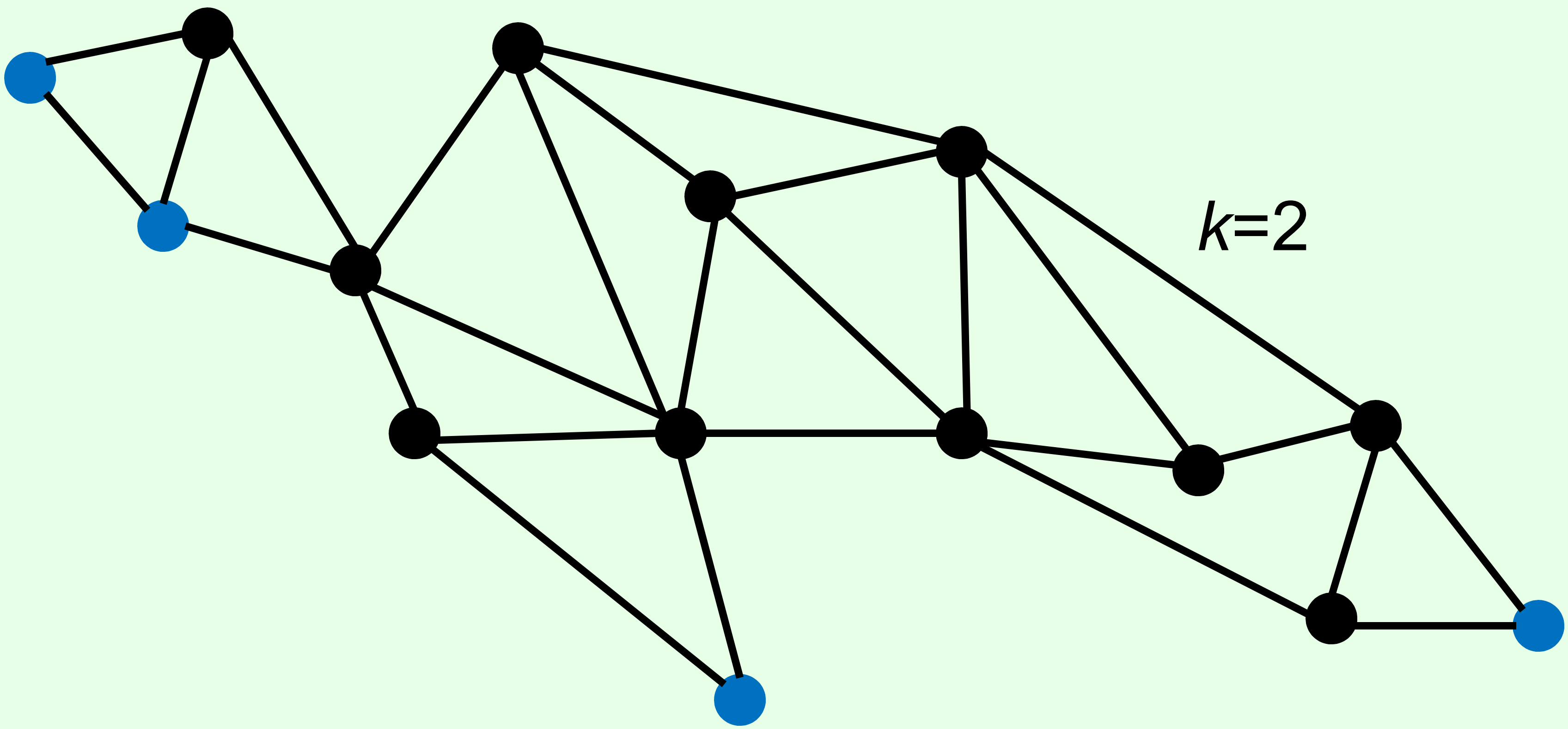
- Xia, C., Lang [2008, 2010, 2011] study rules that do not require CP-nets to be acyclic

Verification

- Instead of starting with trusted nodes, suppose we can actively verify whether nodes are legitimate
  - Nodes that pass the verification step become trusted
- Goal: minimize number of verifications needed so that everyone is deemed legitimate

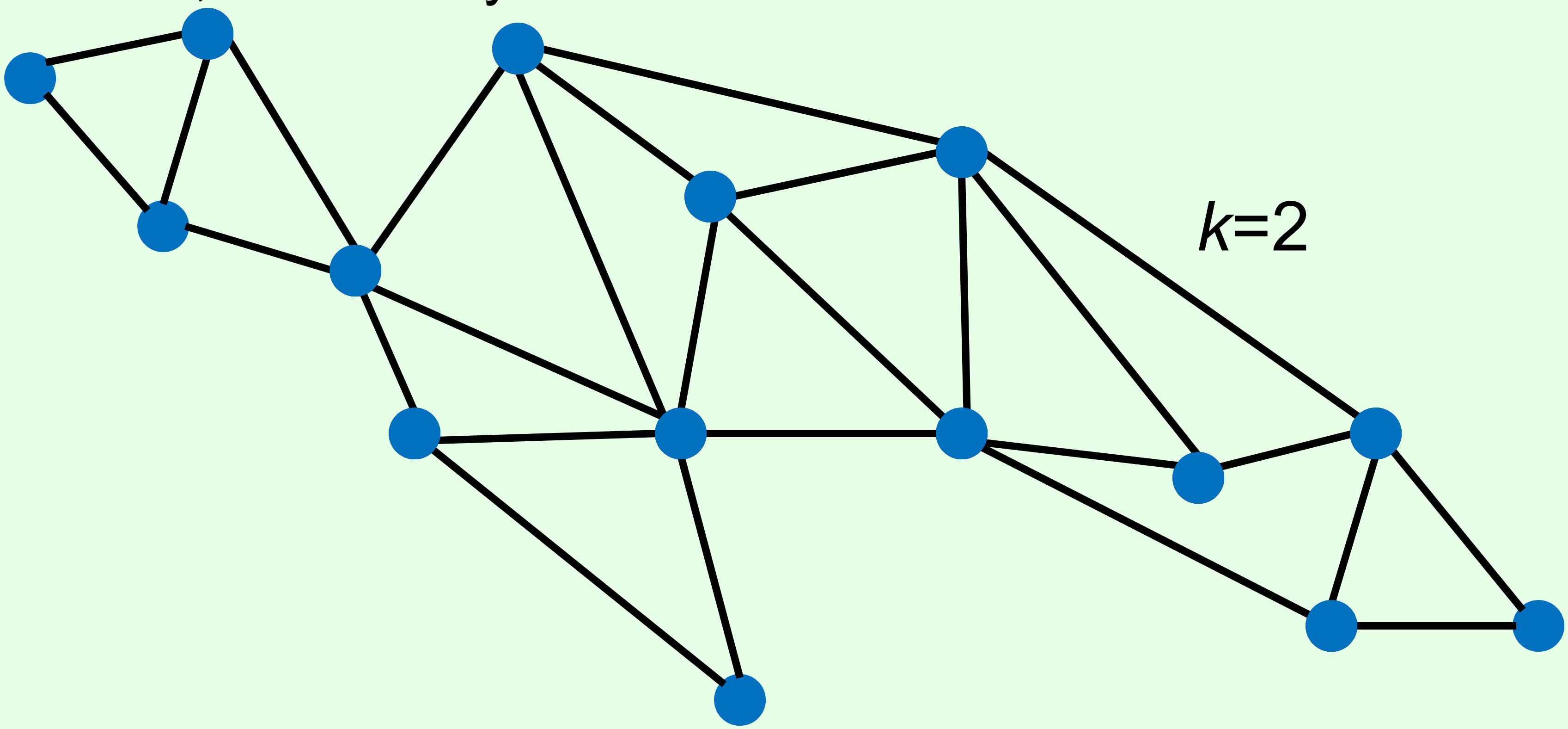
Equivalent to source location problem

- Minimize number of source (=verified) vertices so that nothing is separated from the sources by a vertex cut of at most size *k*
  - I.e. (Menger): there are at least *k+1* vertex-disjoint paths from the sources to each node

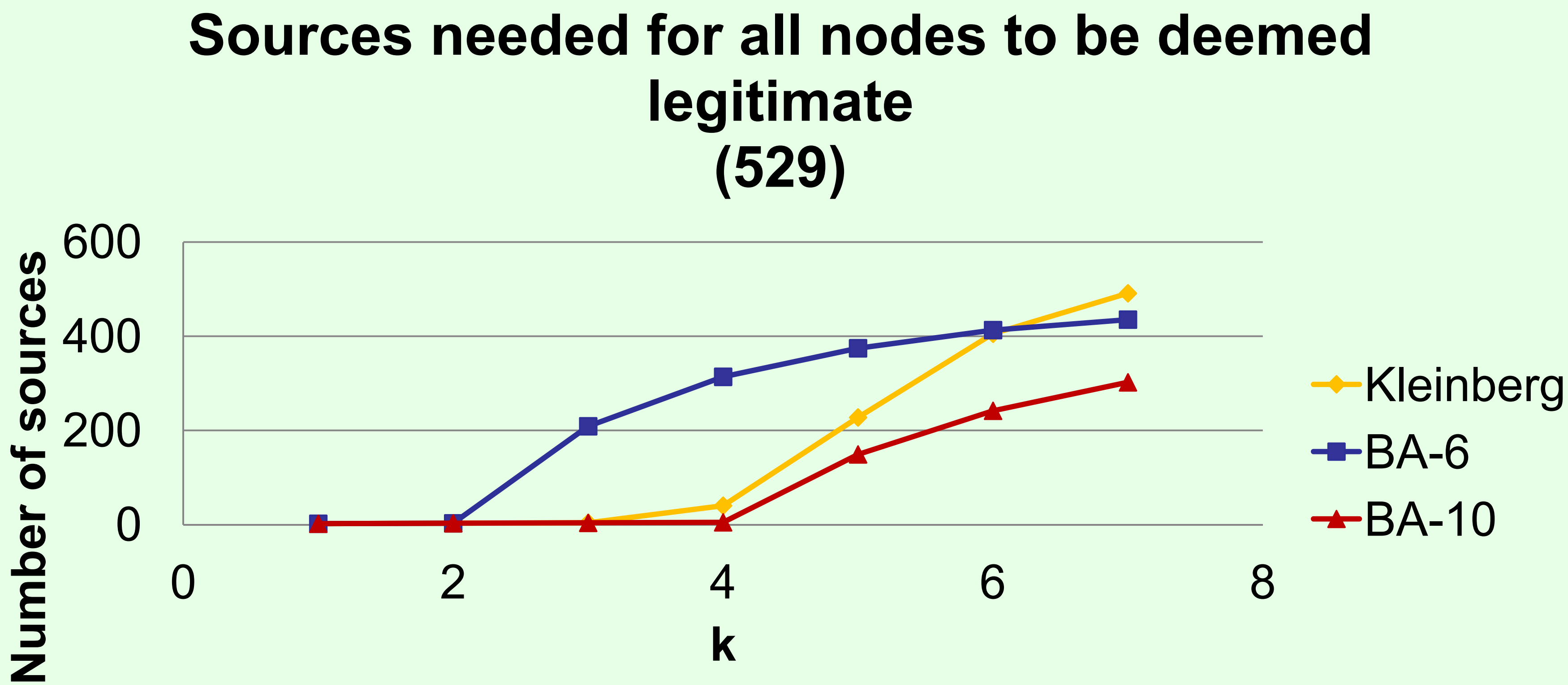


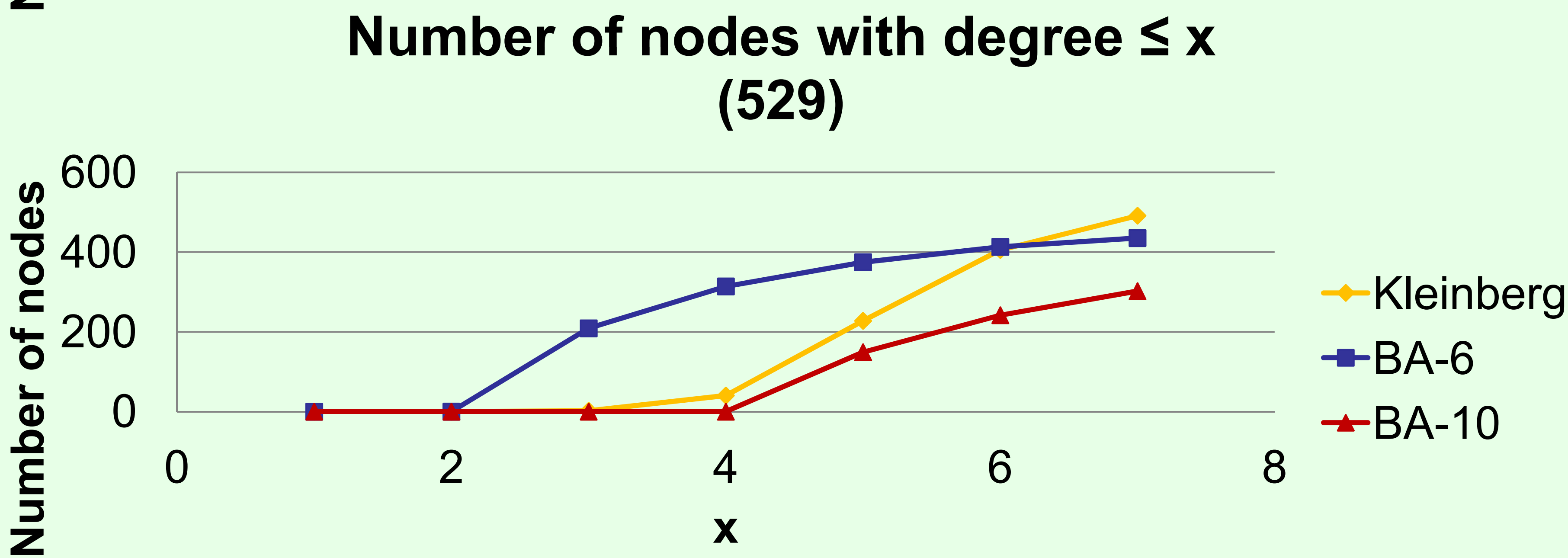
Simple algorithm

- Initial plan: verify everything
- Go through the nodes one by one
  - Check if not verifying that node would make it suspect
  - If not, don't verify it



- Returns an **optimal** solution! (Follows from matroid property [Namagochi, Ishii, Ito 2001])


Sources needed for all nodes to be deemed legitimate
(529)
Number of sources
600
400
200
0
0
2
4
6
8
k
Kleinberg
BA-6
BA-10


Number of nodes with degree ≤ x
(529)
Number of nodes
600
400
200
0
0
2
4
6
8
x
Kleinberg
BA-6
BA-10