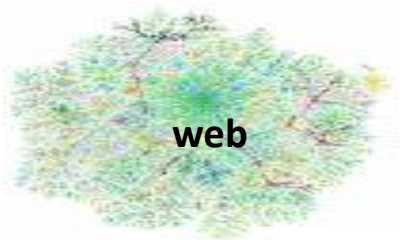
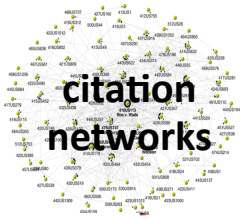
Temporal Analytics in Online Social Networks

Ravi Kumar
Yahoo! Research, Sunnyvale, CA
ravikumar@yahoo-inc.com

Behavioral graphs

- Web graphs
- Host graphs
- Social networks
- Collaboration networks
- Sensor networks
- Biological networks
- …





flickr™

social
networks

LIVEJOURNAL

Research trends

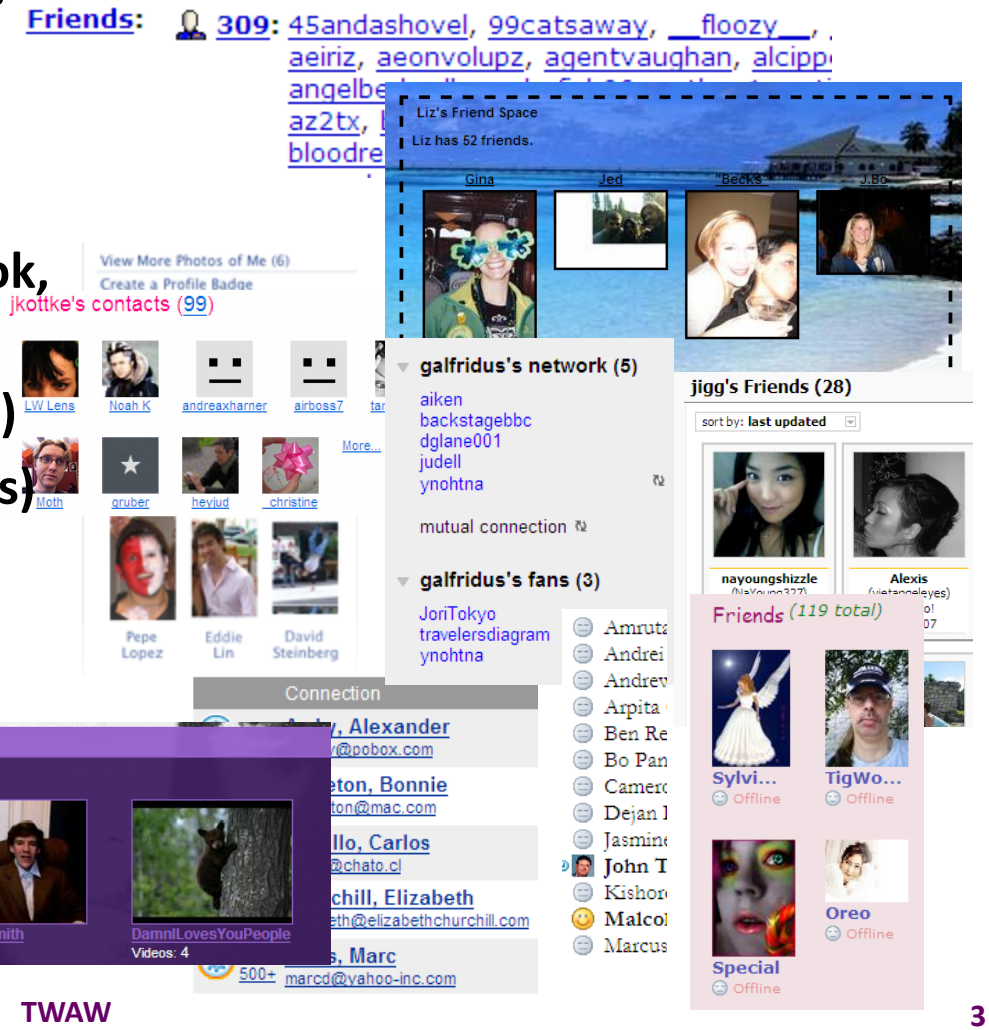
- **Empirical analysis:** examining properties of real-world graphs
- **Modeling:** finding good models for behavioral graphs

There has been a tendency to lump together behavioral graphs arising from a variety of contexts

Social networks

Social network: Graph that represents relationships between users

- **Offline social connections** (friends, clubs, groups, associations)
- **Online social connections** (Facebook, LiveJournal, LinkedIn, Twitter)
- **Messaging** (IM, chat, address book)
- **Social bookmarking** (Digg, Delicious)
- **Content sharing** (Flickr, YouTube)



Studying behavioral graphs

More and more of all interactions are happening online

- The resulting data is a goldmine for studies
  - Massive amounts, even hundreds of millions of nodes
    - Search companies are now working on crawls of 100+ billion pages
    - Facebook has over 600M active users
  - Study phenomena at different scales (eg, interaction of people in focused groups of different sizes, overall structure of the network)
  - Ability to measure, record, and track individual activities at the finest resolution (eg, user befriending another, user buying a dvd, user tagging a photo, user tweeting – when, how, why)
  - Interplay between monadic and dyadic attributes
- A double-edged sword
  - Data is inherently noisy
  - Large scale of data leads to algorithmic challenges
- Graph-theoretic analysis has led to significant impact
  - Link analysis in web search
  - Sophisticated recommendation systems
- Interplay of measurements, modeling, and algorithms

Properties of behavioral graphs

- Degree distributions
  - Heavy tail
- Clustering
  - High clustering coefficient
- Communities and dense subgraphs
  - Abundance; locally dense, globally sparse; spectrum
- Connected components
  - Distribution; "bow-tie" structure
- Connectivity
  - Low diameter; small-world properties
- Compressibility

Analyze these properties over time

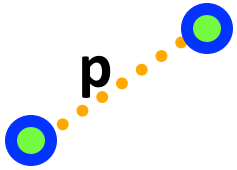
Questions we study

Behavioral networks evolve with additions and deletions of nodes and edges

- **How to model network growth?**
  - **Simple and few parameters**
  - **Consistent with observed phenomena and measurements**
  - **Capture reality as best as possible**
  - **Mathematically tractable**
- **How does network evolve over time?**
  - **Arrivals of nodes and edges**
  - **Change in the graph structure**
  - **Evolution in terms of (graph) properties**
  - **Understand the individual processes behind the evolution**

The G_{np} model

- A random graph model with parameter p
  - Connect each pair of nodes with probability p
  - Erdos-Renyi model, studied almost half-century ago
  - If p is O(c/n), then the graph is sparse

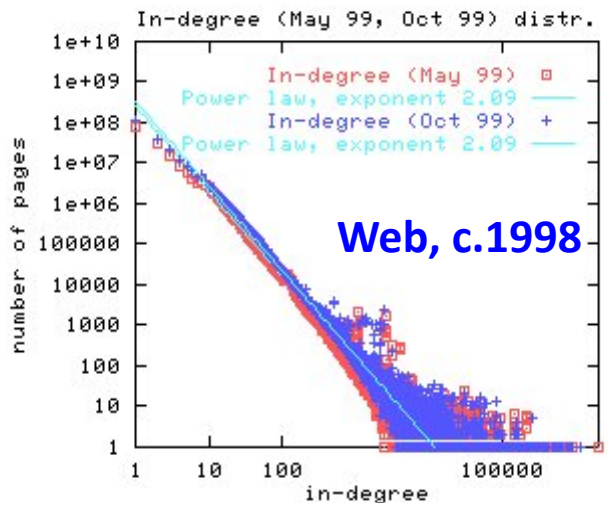


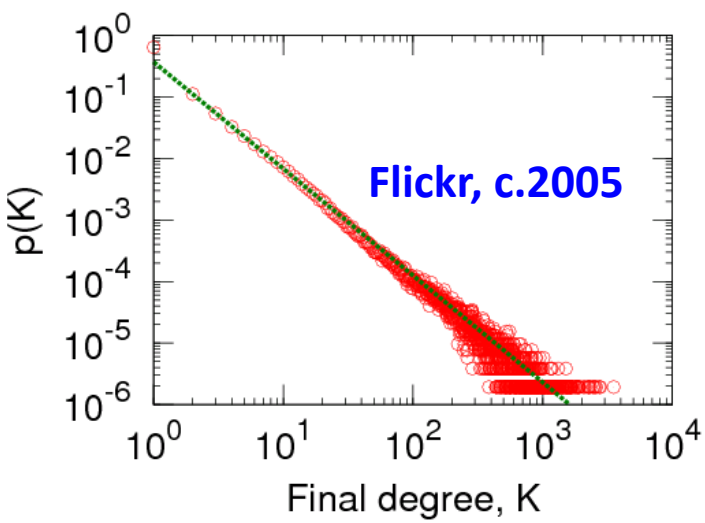
- Many properties are well understood
  - Threshold phenomenon: eg, giant component, connectivity
  - Global properties such as diameter, eigen-values, coloring
  - Probability that a node has degree k is Poisson

$$P(k) = \exp(-\lambda)\, \lambda^k / k!$$

- Evolving version: A new node picks an existing node uniformly at random to add an edge

Power-law degree distributions





- Probability that a node has indegree k, $P(k) \approx 1/k^{\beta}$

  β is the power-law exponent
- Appears straight-line on a log-log plot
- One of the earliest empirical observations about web and social networks
- Hence, G_{np} is not an appropriate model for social networks

Preferential attachment model

Barabasi, Albert, Jeong 1999

- Observation: Rich-get-richer
  - Popular documents get more citations
  - Popular people get more friends
- Each step has one new incoming node along with an edge
- Probability it links to another node is proportional to how popular is the latter, ie, its degree

$$\Pr[\text{new node links to node } i] = d_i / \sum d_j$$

Theorem. In this model, power-law exponent = 3

Intuitive proof. $\partial d_i / \partial t = d_i / (2t)$

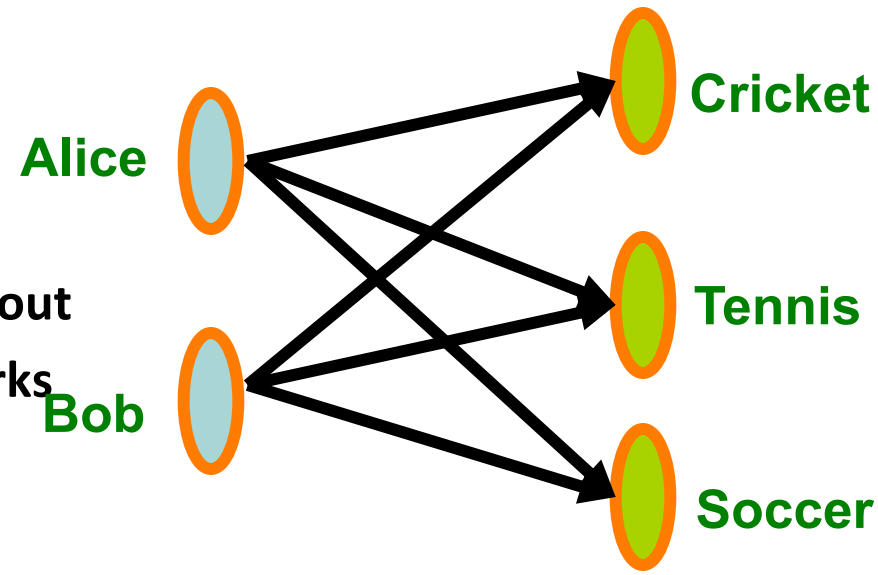
If node i was added at time t_i, then $d_i(t) = (t/t_i)^{0.5}$

$\Pr[d_i(t) > k] = \Pr[t_i < t/k^2] = 1/k^2$

Communities

Edges usually imply endorsement or interest in a topic or a person

- Users link to pages they care about
- Friendship links in social networks
- Two users with similar interests need not know each other



Communities are dense subgraphs or dense bipartite subgraphs

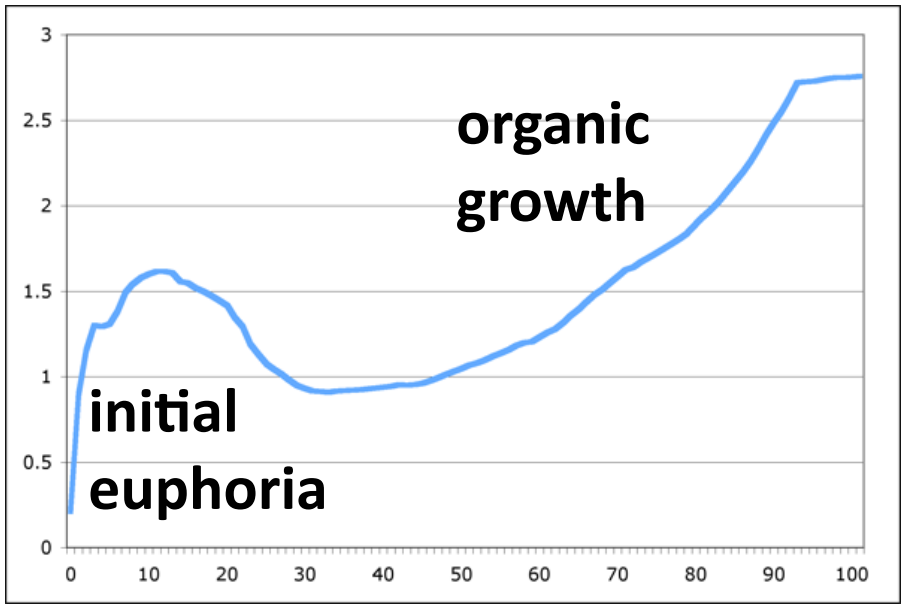
Web and social networks are abundant In communities

Copying model

Kumar et al 2000

- **Observation:** People copy their friend's webpage when creating a new one or copy their friend's contacts when joining a social network
- When a new node arrives, it copies edges from a pre-existing node with probability $1 - \alpha$
  - links to the destination of the edge
- The degree distribution is a power-law with exponent $(2 - \alpha)/(1 - \alpha)$
- Can explain communities: The number of dense bipartite cliques in this model is large

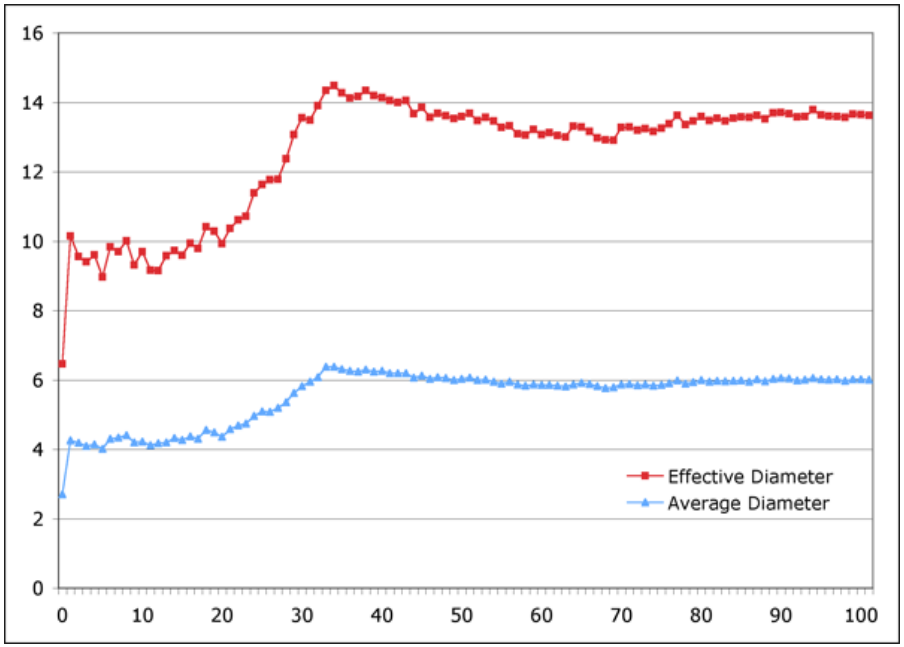
Flickr: Density, diameter over time



Density

Shrinking diameters and densification in citation graphs:
Leskovec, Kleinberg, Faloutsos
2005

Diameter



Forest-fire model

Leskovec, Kleinberg, Faloutsos 2005

Observation: Copying happens beyond one step

When a new node arrives, it

- copies an edge from a pre-existing node with prob. 1 - α
- copies an edge from the destination of the edge
- …

An iterated version of the copying model

In addition to the above, leads to densification and shrinking diameters, in empirical simulations

Affiliation networks model

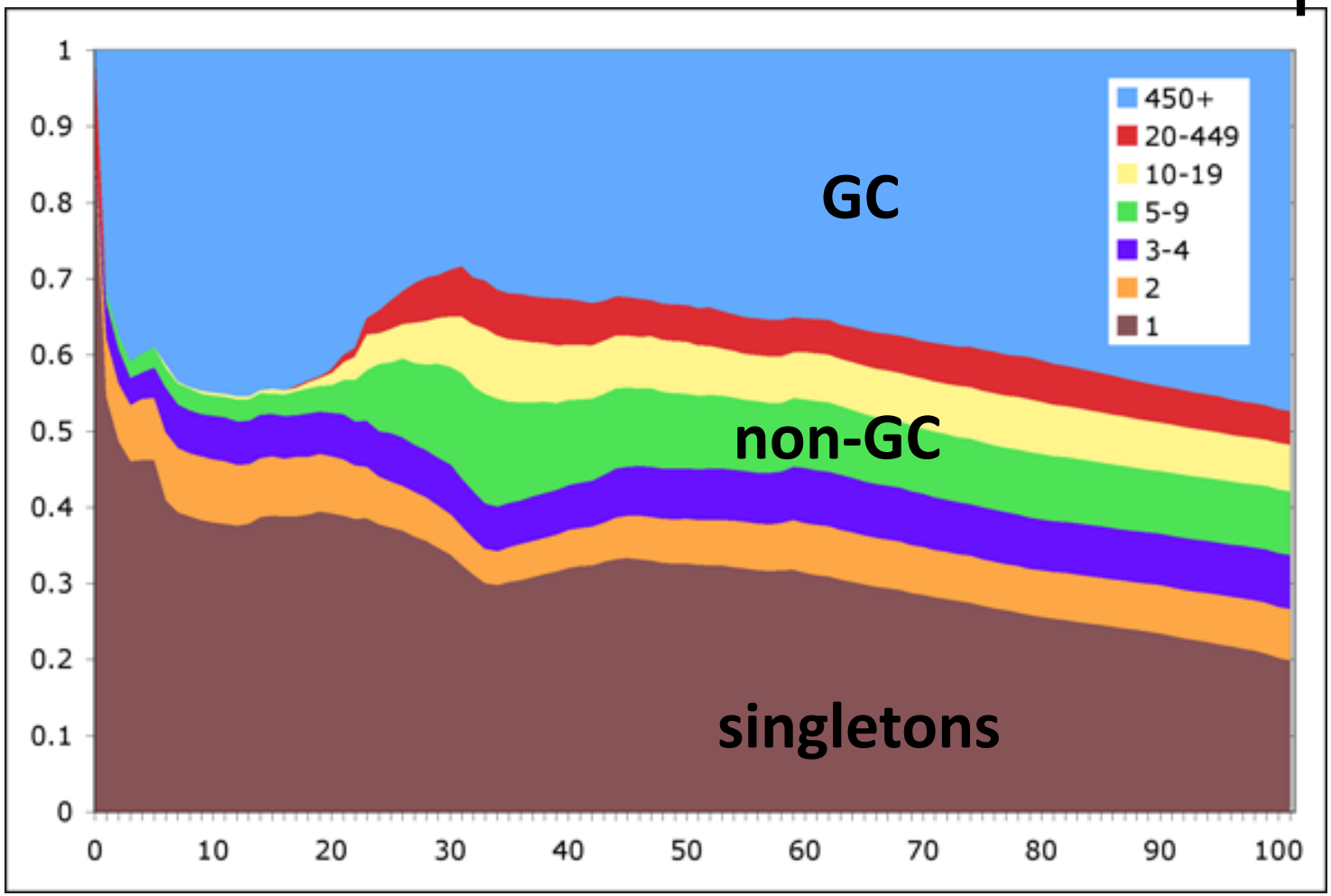
- Bipartite graph B(Q, U, D) and a graph G(Q, E)
  - Q = papers, U = topics
- Co-evolution of B and G
  - Q side of B evolves by copying
  - U side of B evolves by copying
  - Q side of G evolves by prototyping (via evolution of Q and U in B)
- This evolutionary model produces graphs with densification and shrinking diameters Lattanzi, Sivakumar 2009

Dynamic analysis via snapshots

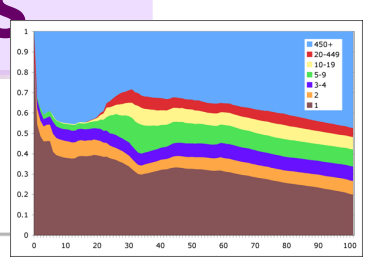
- Evolution of web pages and content
  - Fetterly, Manasse, Najork, Wiener 2004
  - Ntoulas, Cho, Olston 2004
  - Adar, Teevan, Dumais, Elsas 2009
- Social network evolution
  - Kossinetts, Watts 2006
- Citation graphs, autonomous systems
  - Katz 2005
  - Leskovec, Kleinberg, Faloutsos 2005
  - Densification and shrinking diameters
- Process dynamics
  - Backstrom, Huttenlocher, Kleinberg, Lan 2006
  - Leskovec, Adamic, Huberman 2006

Components: A grand canyon view

Flickr



Dynamics of the components



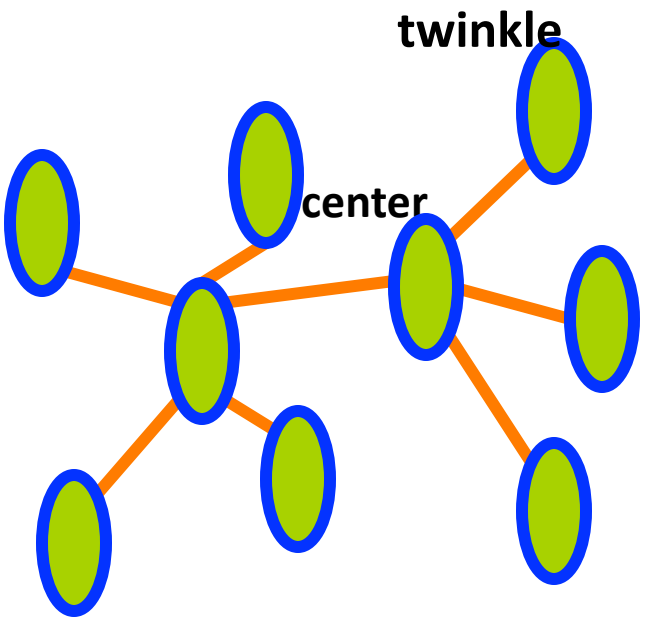
- How do components consolidate?

	1	2	3-4	5-9	10-19	20-449	450+
1	205.1						
2	55.9	0.8					
3-4	64.2	0.5	0.3				
5-9	70.8	0.4	0.3	0.2			
10-19	43.9	0.2	0.1	0.1	0.09		
20-449	2.6	0.1	0.01	0.07	0.04	0.03	
450+	315.3	11.5	7.1	5.0	2.4	1.0	0

  - Singletons merge with non-GC and GC
  - Non-GCs merge with GC
  - Almost never a non-GC merges with another non-GC
- Why is singleton attracted to a non-GC?
  - Is there a special attractor in a non-GC?

Structure of non-GC: Stars

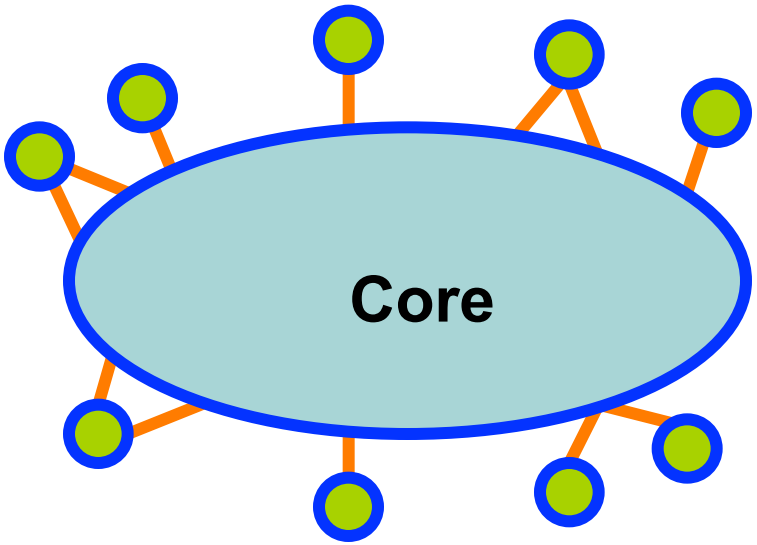
- Informal definition of star
  - There are one or more centers (high degree nodes)
  - There are many degree-one nodes (twinkles) connected to these centers

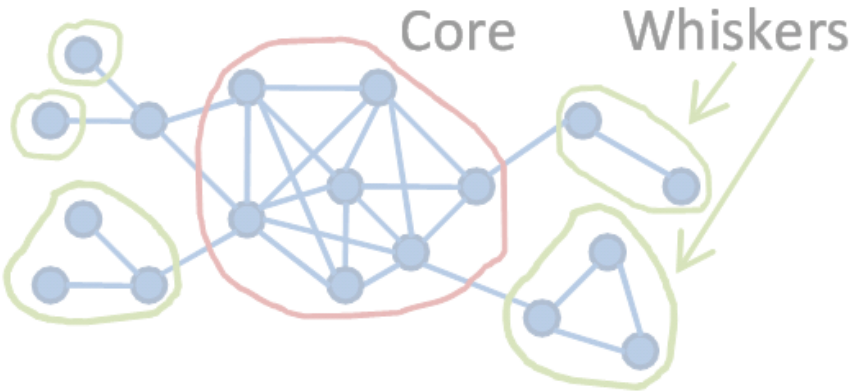


- Under reasonable setting of parameters, more than 93% of non-GCs are stars
- The stars form quickly
- A large fraction of them are yet to be absorbed into GC

Structure of GC: Core

- There is a small core of very high connectivity inside GC
- The core is not comprised of star centers
- GC connectivity does not depend on star centers





This has implications for finding dense communities: Leskovec, Lang, Dasgupta, Mahoney 2008

A simple model with user types

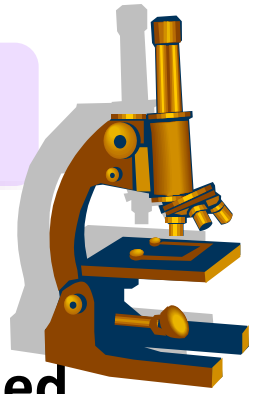
At each time step

- A person joins the network and is chosen to be one of three types: passive user, inviter, linker
- Few friendships (ie, edges) arrive
  - Source of edge chosen from inviters/linkers with degree-biased probabilities (ie, preferential attachment)
  - If source = inviter, destination = a new passive user
  - If source = linker, destination chosen from linkers and inviters, degree-biased

Empirically, this model generates the observed temporal characteristics (fraction of components, stars, core)

Open question. Can we say anything formal?

A microscopic look at evolution



- We are starting to have more nuanced models
  - Processes such as preferential attachment are assumed without actually being observed
  - There are many ways to generate power-law degrees
- What is the best way to compare two graph models?
  - Maximum-likelihood (standard tool in ML)
  - Efficiency issues: Bezakova, Kalai, Santhanam 2006
- We have edge-by-edge arrival information, so can take an edge and compare the likelihood of its existence in competing models

How does the network evolve?

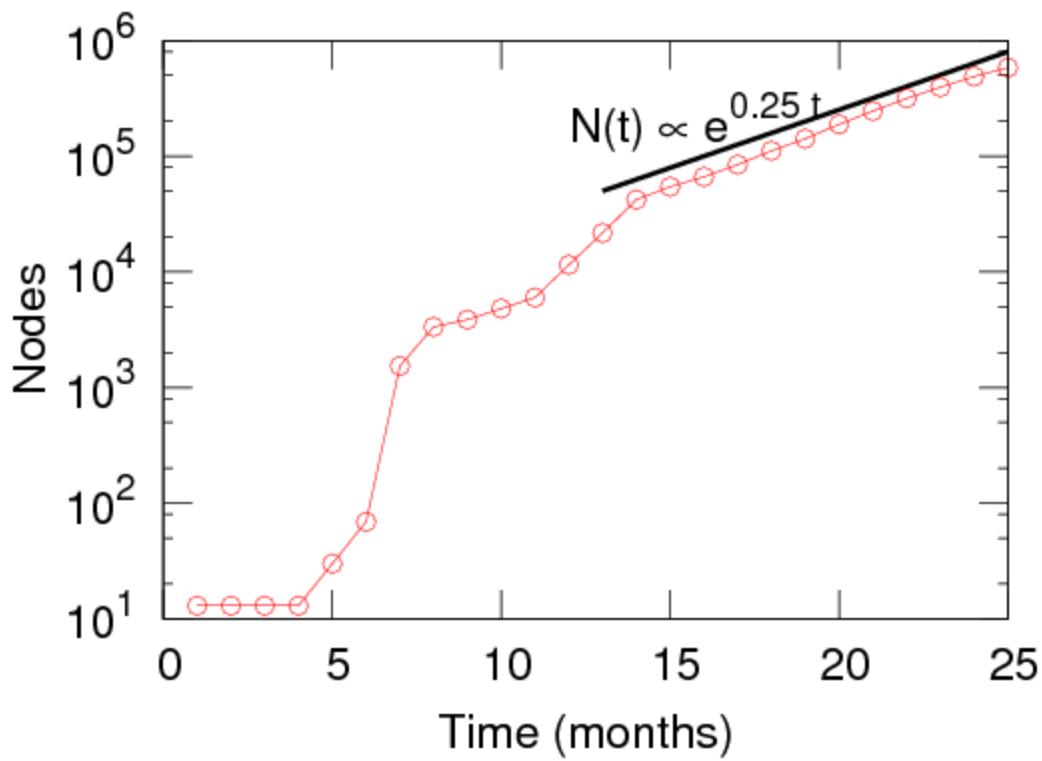
Three processes govern the evolution

1. **Node arrival process.** Nodes enter the network
2. **Edge initiation process.** Each node decides when to initiate an edge
3. **Edge destination process.** Determines destination after a node decides to initiate

We will present a complete model of network evolution by mining the node and edge creation data

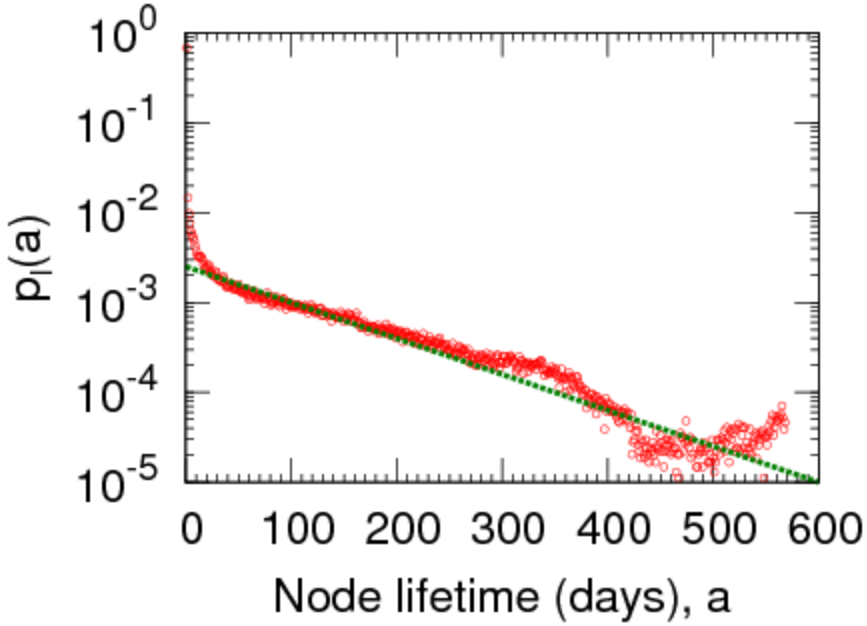
Node arrivals: Rate

It really depends on the network, ranging from sub-linear to exponential



Node arrivals: Lifetime

Lifetime a: time between node's first and last edge

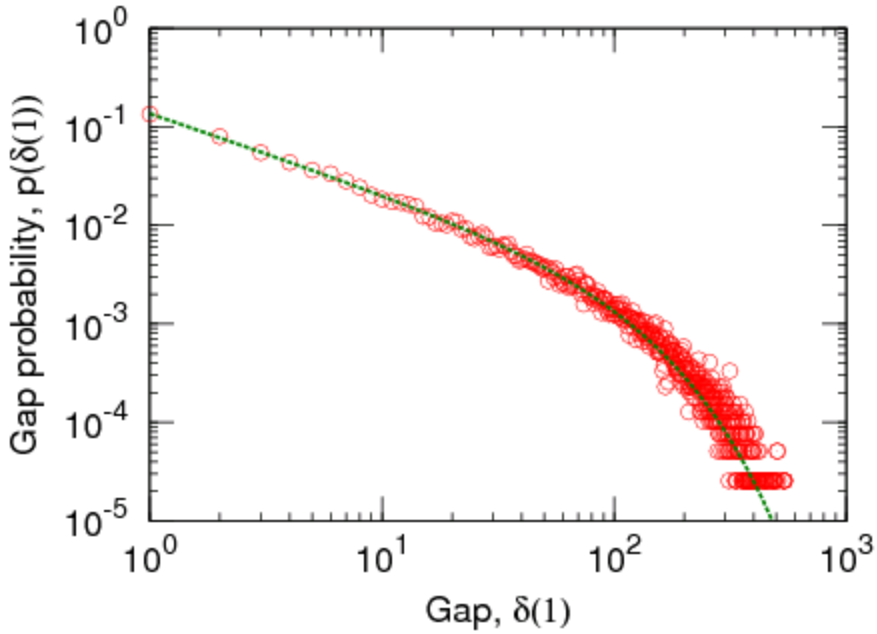


Node lifetime is exponentially distributed:

$$p_l(a) = \lambda \exp(-\lambda a)$$

How are edges initiated?

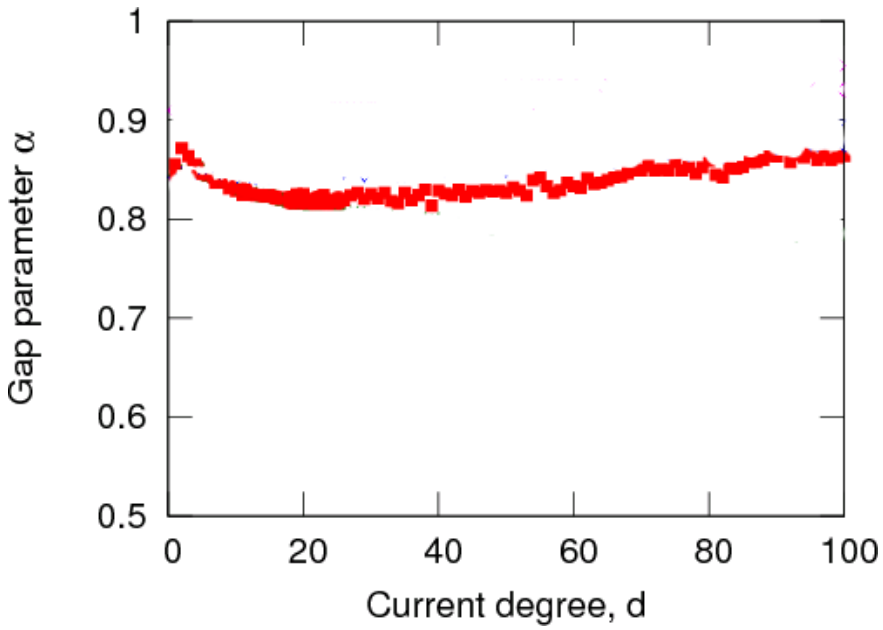
Edge gap $\delta(d)$: time between d^{th} and $d+1^{st}$ edge

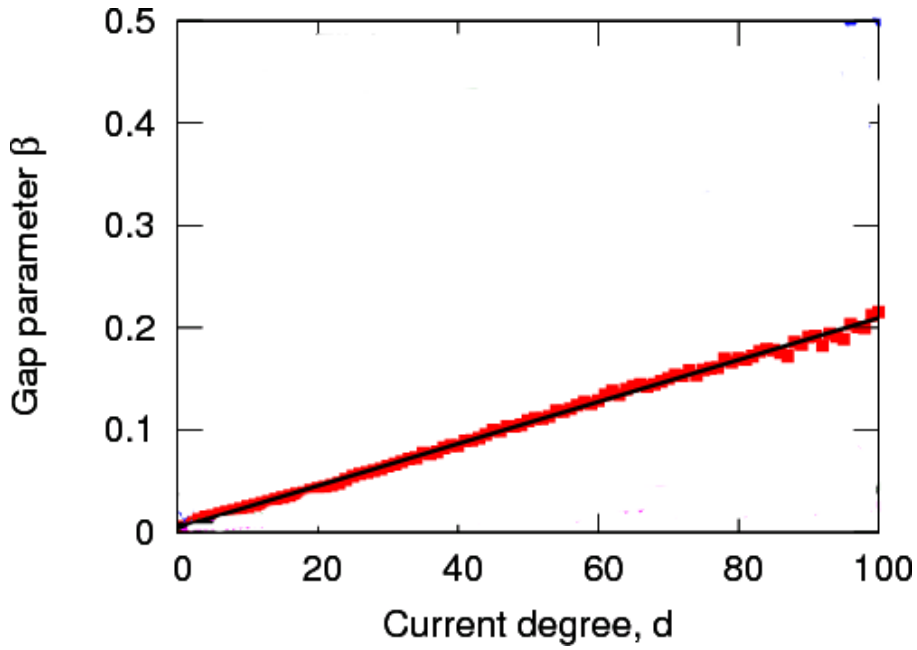


Competing models: exponential, lognormal, stretched exponential, power-law with exponential cutoff

Edge interarrivals follow power-law with exponential cutoff: $p_g(\delta(d); \alpha(d), \beta(d)) \propto \delta(d)^{-\alpha(d)} \exp(-\beta(d)\, \delta(d))$

How do α and β change with degree?





α(d) (power-law part) is constant

β(d) (exp-cutoff part) is linear in d

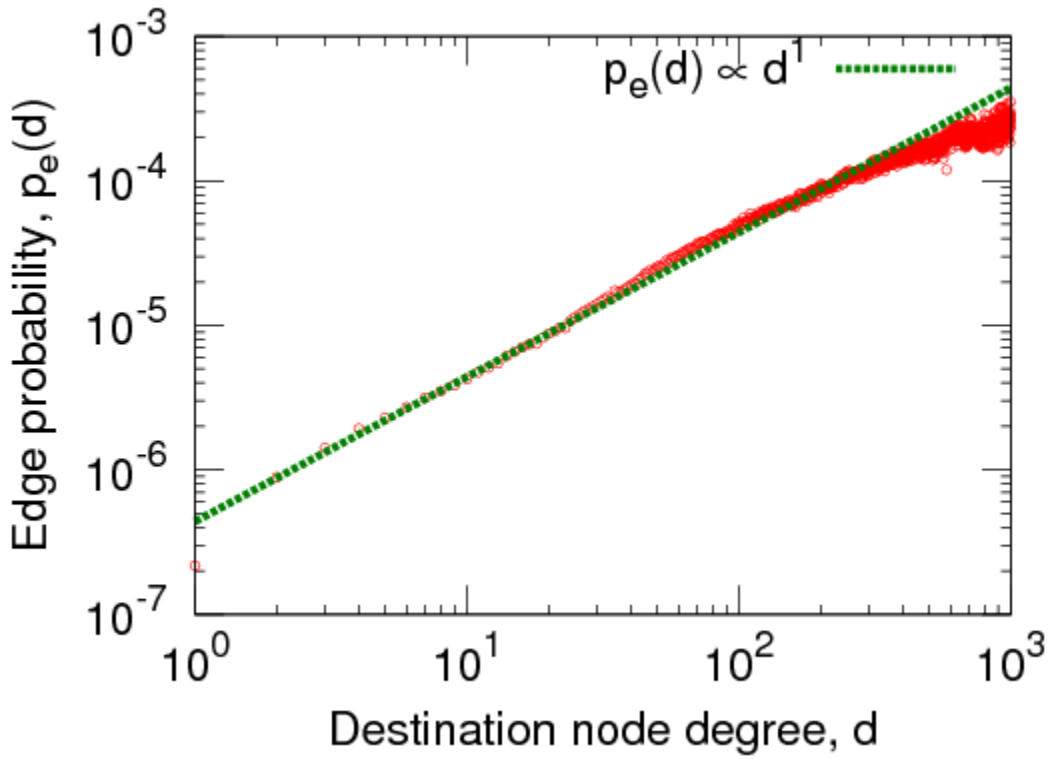
This means nodes of higher degree start adding edges faster and faster

What we know so far

Process	Model
1) Node arrival	• Node arrival is network dependent • Node lifetime: $p_l(a) = \lambda \exp(-\lambda a)$
2) Edge initiation	• Edge gaps: $p_g(\delta) \propto \delta^{-\alpha} \exp(-\beta d\ \delta)$
3) Edge destination	

Does preferential attachment happen?

We unroll the true network edge arrivals and measure node degrees where edges attach



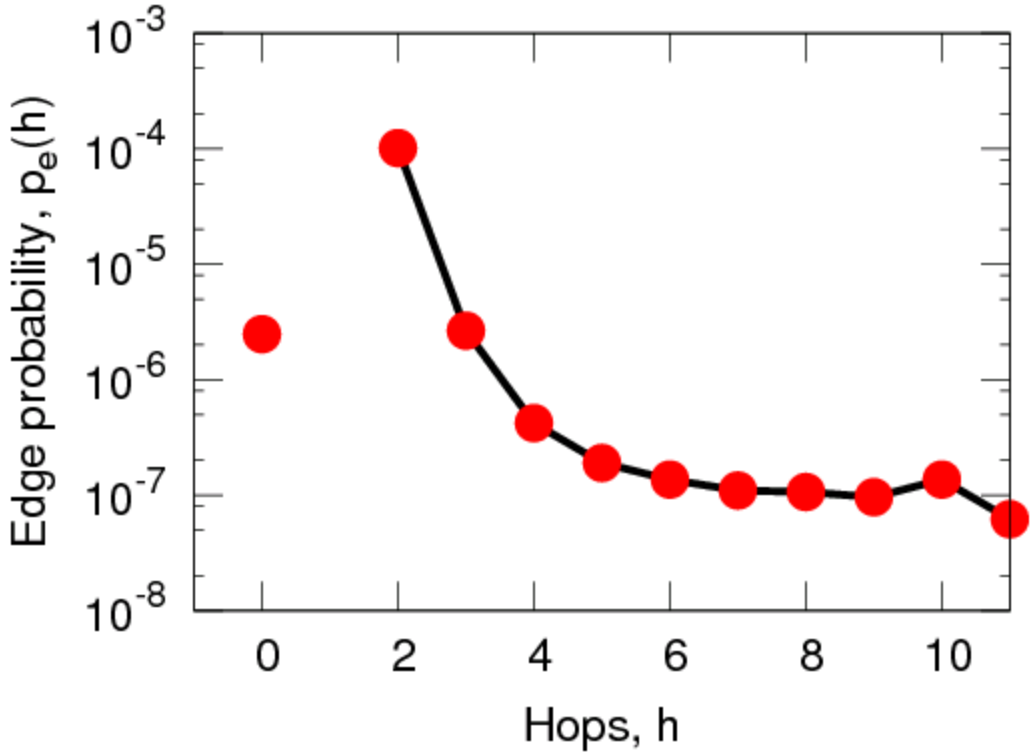
First explicit proof that preferential attachment indeed happens

But, hold on …

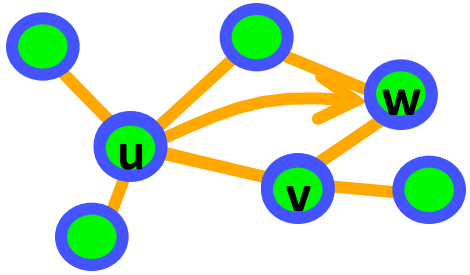
How local are the added edges?

Just before edge (u,v) is placed, how far are u and v?

Normalize this by number of nodes at that hop distance



Real edges are local and most of them (66%) are triangle closing



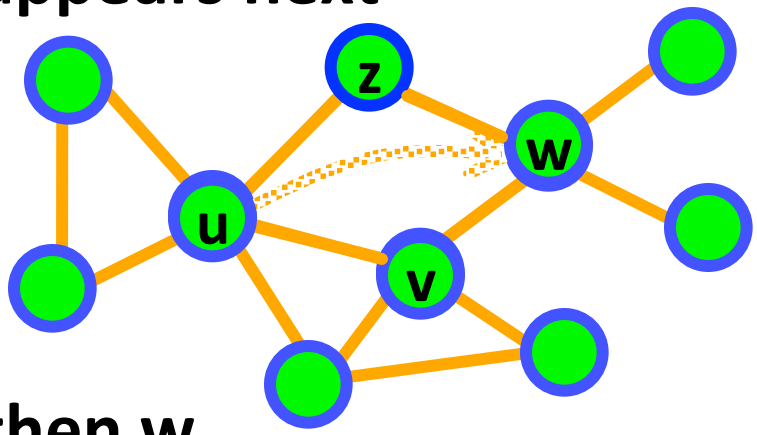
Long known to sociologists

George Simmel (1858-1918),

Krackhardt and Handcock 2007

Closing triangles

- New triangle-closing edge (u,w) appears next
- We model this as
  1. Choose u's neighbor v
  2. Choose v's neighbor w
  3. Add edge (u,w)



- 25 strategies for choosing v and then w
  - Random, degree preferentially, number of common friends, time of last activity, combination
- Can compute likelihood of each strategy
  - Under random-random: p(u,w) = 1/5*1/2+1/5*1/4

Triangle closing strategies

Log-likelihood improvement over the baseline

Strategy to select v (1st node)

Select w (2nd node)

FLICKR	random	$deg^{0.2}$	com	$last^{-0.4}$	$comlast^{-0.4}$
random	13.6	13.9	14.3	16.1	15.7
$deg^{0.1}$	13.5	14.2	13.7	16.0	15.6
$last^{0.2}$	14.7	15.6	15.0	17.2	**16.9**
com	11.2	11.6	11.9	13.9	13.4
$comlast^{0.1}$	11.0	11.4	11.7	13.6	13.2

Strategies to pick a neighbor

- random: uniformly at random
- deg: proportional to its degree
- com: prop. to the number of common friends
- last: prop. to time since last activity
- comlast: prop. to com*last

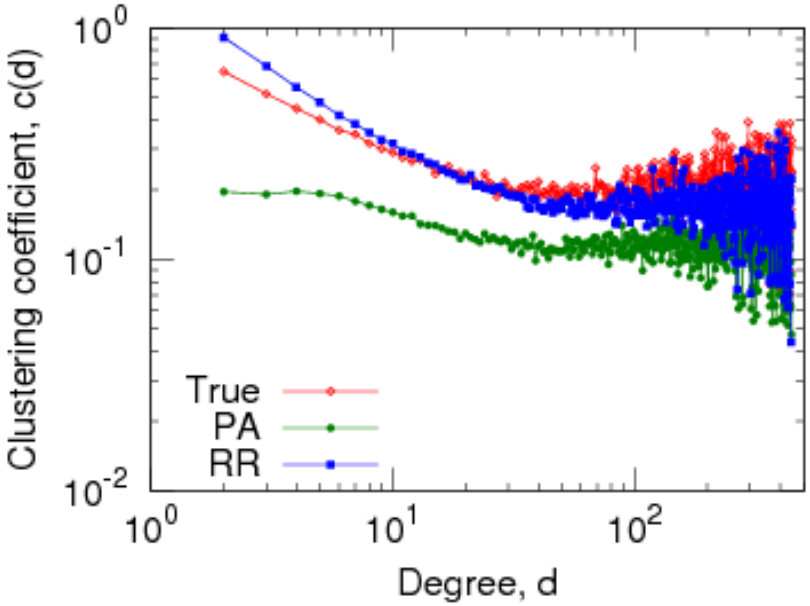
Random-random works quite well

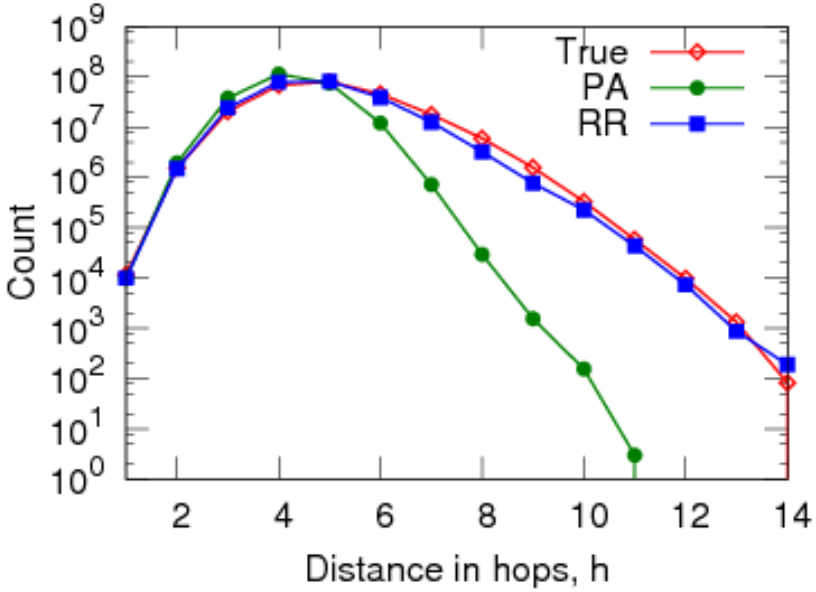
What we know so far

Process	Model
1) Node arrival	• Node arrival is network dependent • Node lifetime: $p(a) = \lambda \exp(-\lambda a)$
2) Edge initiation	• Edge gaps: $p(\delta) \propto \delta^{-\alpha} \exp(-\beta d\ \delta)$
3) Edge destination	• First edge chosen preferentially • Use random-random strategy to close triangles

Semi-simulation: Closer to truth

Take the network at T/2 and evolve it using preferential attachment (PA) and random-random (RR) for edge addition events





The complete network model

- Nodes arrive using the arrival rate
- Node u arrives
  - It has lifetime $a \sim \lambda \exp(-\lambda a)$
  - Adds first edge to node v with probability proportional to degree of v
- A node u with degree d has gap $\delta \sim \delta^{-\alpha} \exp(-\beta d\, \delta)$ and goes to sleep for δ time steps
- When u wakes up, if its lifetime still valid, creates a random-random triangle-closing edge

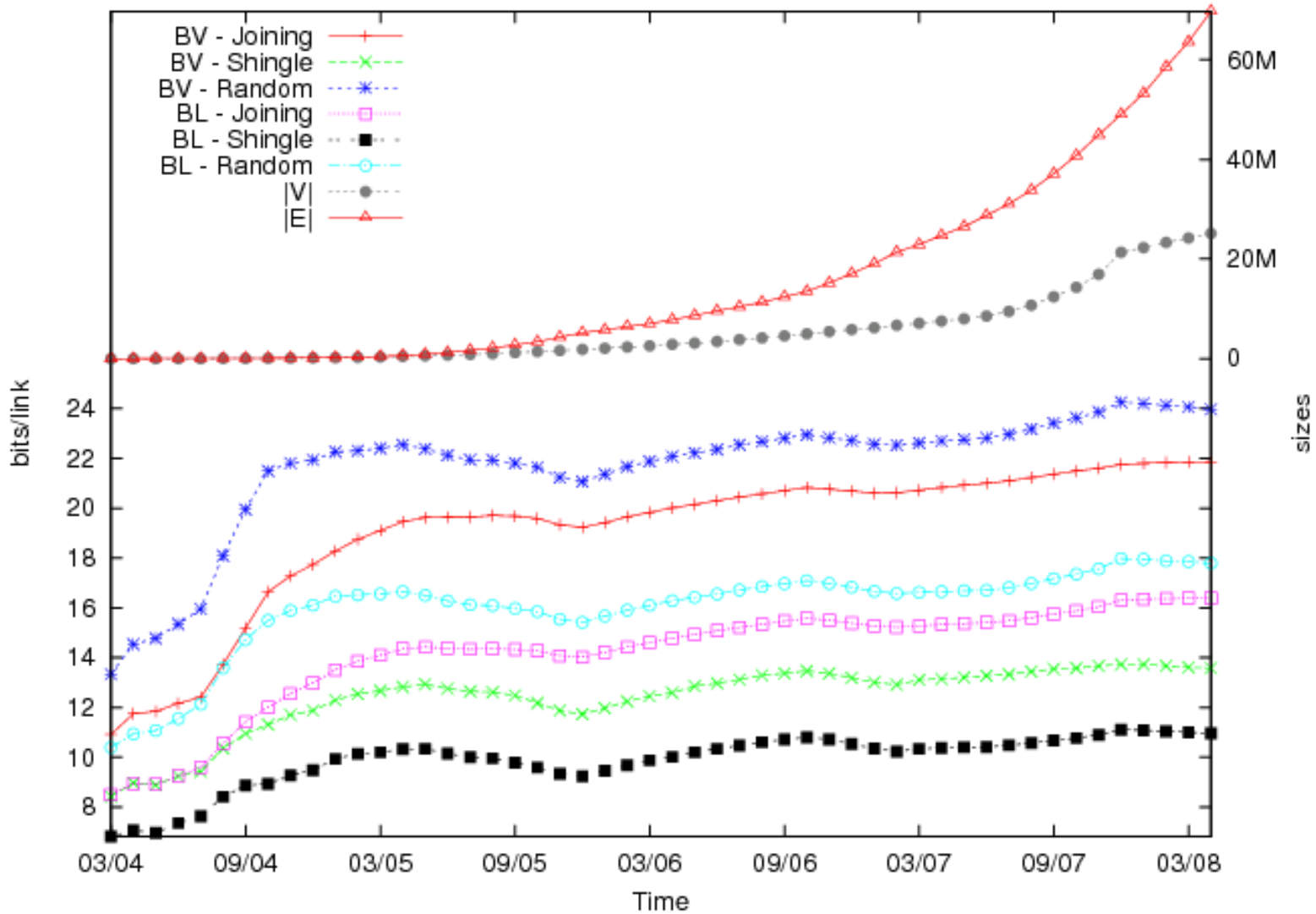
An analysis

Theorem. The out-degrees are distributed according to a power-law with exponent

$$1 + (\lambda / \beta) \cdot (\Gamma(2-\alpha) / \Gamma(1-\alpha))$$

- For Flickr, true exponent = 1.73, $\lambda = 0.0092$, $\alpha = 0.84$, $\beta = 0.002$, calculated exponent = 1.74
- Analogous results hold for del.icio.us, Yahoo! Answers, LinkedIn
- Interesting as temporal behavior leads to power-law degree distribution

Compressibility of Flickr over time



Compressibility of the Web

- **Snapshots of the web graph can be losslessly compressed using less than 3 bits per edge**

  Boldi, Vigna WWW 2004

- **Improved to ~2 bits using another data mining-inspired compression technique**

  Buehrer, Chellapilla WSDM 2008

- **Subsequent improvements**

  Boldi, Santini, Vigna WAW 2009

18.5 Mpages, 300 Mlinks from .uk									
R	Average reference chain			Bits/node			Bits/link		
	$W = 1$	$W = 3$	$W = 7$	$W = 1$	$W = 3$	$W = 7$	$W = 1$	$W = 3$	$W = 7$
∞	171.45	198.68	195.98	44.22	38.28	35.81	2.75	2.38	2.22
3	1.04	1.41	1.70	62.31	52.37	48.30	3.87	3.25	3.00
1	0.36	0.55	0.64	81.24	62.96	55.69	5.05	3.91	3.46
Tranpose									
∞	18.50	25.34	26.61	36.23	33.48	31.88	2.25	2.08	1.98
3	0.69	1.01	1.23	37.68	35.09	33.81	2.34	2.18	2.10
1	0.27	0.43	0.51	39.83	36.97	35.69	2.47	2.30	2.22
118 Mpages, 1 Glinks from WebBase									
R	Average reference chain			Bits/node			Bits/link		
	$W = 1$	$W = 3$	$W = 7$	$W = 1$	$W = 3$	$W = 7$	$W = 1$	$W = 3$	$W = 7$
∞	85.27	118.56	119.65	30.99	27.79	26.57	3.59	3.22	3.08
3	0.79	1.10	1.32	38.46	33.86	32.29	4.46	3.92	3.74
1	0.28	0.43	0.51	46.63	38.80	36.02	5.40	4.49	4.17
Tranpose									
∞	27.49	30.69	31.60	27.86	25.97	24.96	3.23	3.01	2.89
3	0.76	1.09	1.31	29.20	27.40	26.75	3.38	3.17	3.10
1	0.29	0.46	0.54	31.09	29.00	28.35	3.60	3.36	3.28

Key insights

1. Many web pages have similar set of neighbors

2. Edges tend to be "local"

Main ideas of Boldi-Vigna

Canonical ordering: Sort URLs lexicographically, treating them as strings Randall et al 2002

…
17: www.uchicago.edu/alchemy
18: www.uchicago.edu/biology
19: www.uchicago.edu/biology/plant
20: www.uchicago.edu/biology/plant/copyright
21: www.uchicago.edu/biology/plant/people
22: www.uchicago.edu/chemistry
…

This gives an identifier for each URL

Source and destination of edges are likely to get nearby IDs

- Templated webpages
- Many edges are intra-host or intra-site

- Due to templates, the adjacency list of a node is similar to one of the 7 preceding URLs in the lexicographic ordering
- Express adjacency list in terms of one of these
- Eg, consider these adjacency lists

1: 1, 2, 4, 8, 16, 32, 64
2: 1, 4, 9, 16, 25, 36, 49, 64
3: 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144
4: 1, 4, 8, 16, 25, 36, 49, 64

Encode as (-2), remove 9, add 8

BV compression algorithm

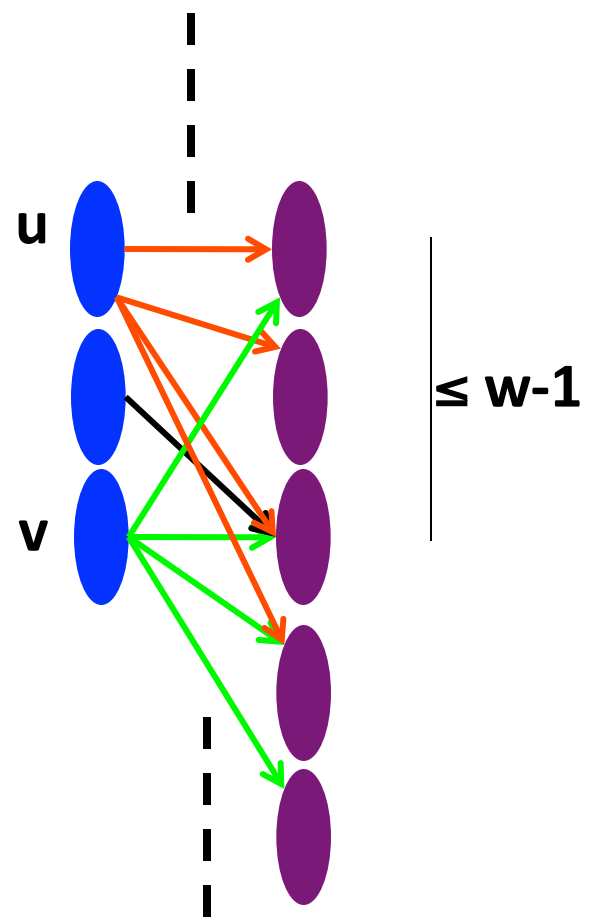
Each node has a unique ID from the canonical ordering

Let w = copying window parameter

To encode a node v

- Check if out-neighbors of v are similar to any of w-1 previous nodes in the ordering
- If yes, let u be the leader: use lg w bits to encode the gap from v to u + difference between out-neighbors of u and v
- If no, write lg w zeros and encode out-neighbors of v explicitly

Use gap encoding on top of this



Canonical orderings

- BV compressions depend on a canonical ordering of nodes
  - This canonical ordering should exploit neighborhood similarity and edge locality
- How do we get a good canonical ordering?
  - Unlike the web page case, it is unclear if social networks have a natural canonical ordering
- Caveat: BV is only one genre of compression scheme
  - Lack of good canonical ordering does not mean graph is incompressible

Some natural canonical orderings

- Random order
- Natural order
  - Time of joining in a social network
  - Lexicographic order of URLs
  - Crawl order
- Graph traversal orders
  - BFS and DFS
- Use attributes of the nodes
  - Eg, Geographic location: order by zip codes
  - May produce a bucket order
- Ties can be broken using more than one order

Shingle ordering heuristic

- Obtain a canonical ordering by bringing nodes with similar neighborhoods close together
- Fingerprint neighborhood of each node
  - Order the nodes according to the fingerprint
  - If fingerprint can capture neighborhood similarity and edge locality, then it will produce good compression via BV, provided the graph is amenable
- Use Jaccard coefficient to measure similarity between nodes
  - $J(A, B) = |A \cap B| / |A \cup B|$
- Double shingle order: break ties within shingle order using a second shingle

Performance of shingle ordering

BV

Graph	Natural	Shingle	Double shingle
Flickr	21.8	13.5	13.5
UK host	10.8	8.2	8.1
IndoChina	2.02	2.7	2.7

BV + reciprocal links

Graph	Natural	Shingle	Double shingle
Flickr	16.4	10.9	10.9
UK host	10.5	8.2	8.1
IndoChina	2.35	2.7	2.7

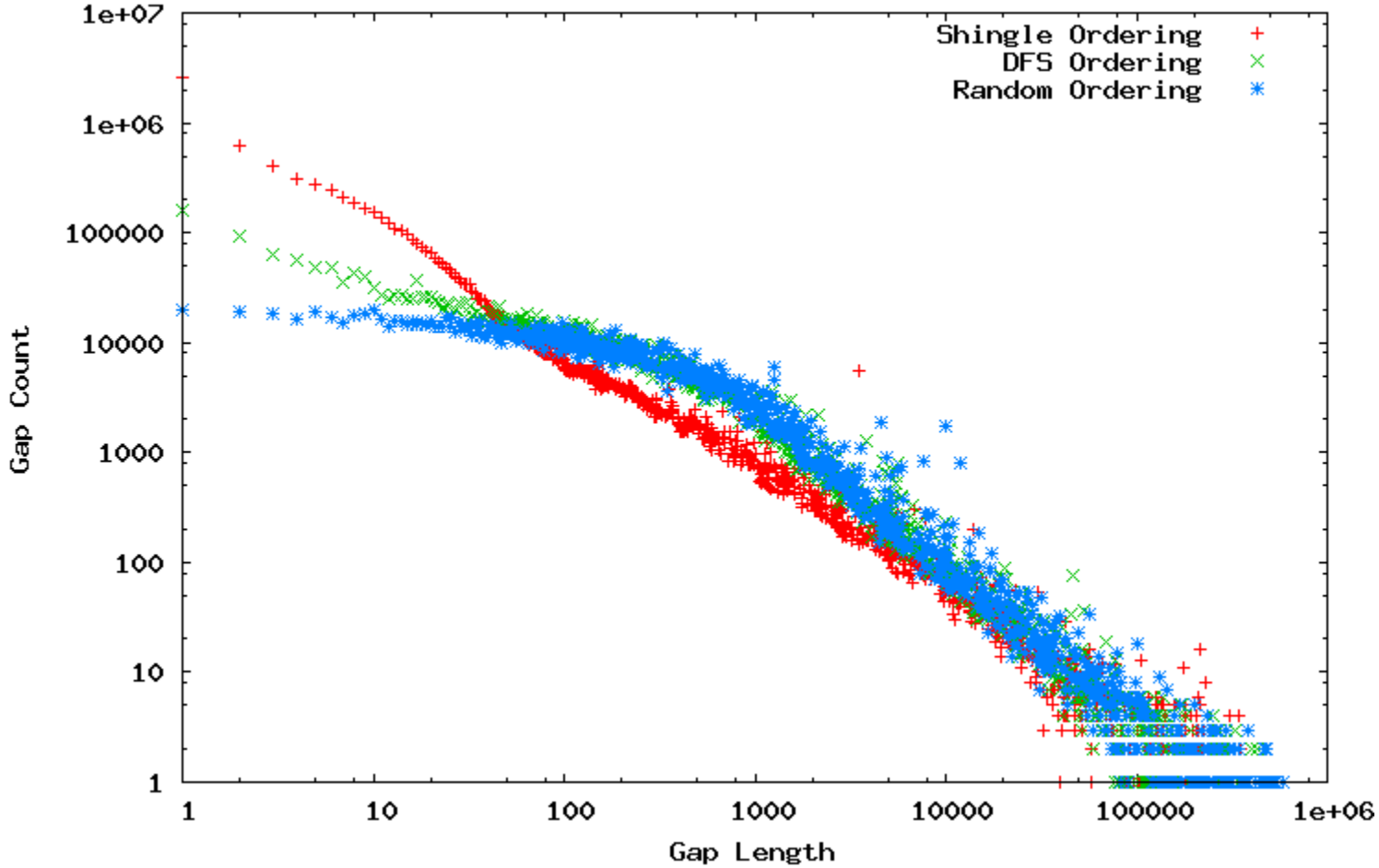
Geography does not seem to help for Flickr graph

A property of shingle ordering

Theorem. Using shingle ordering, a constant fraction of edges will be "copied" in graphs generated by preferential attachment/copying models

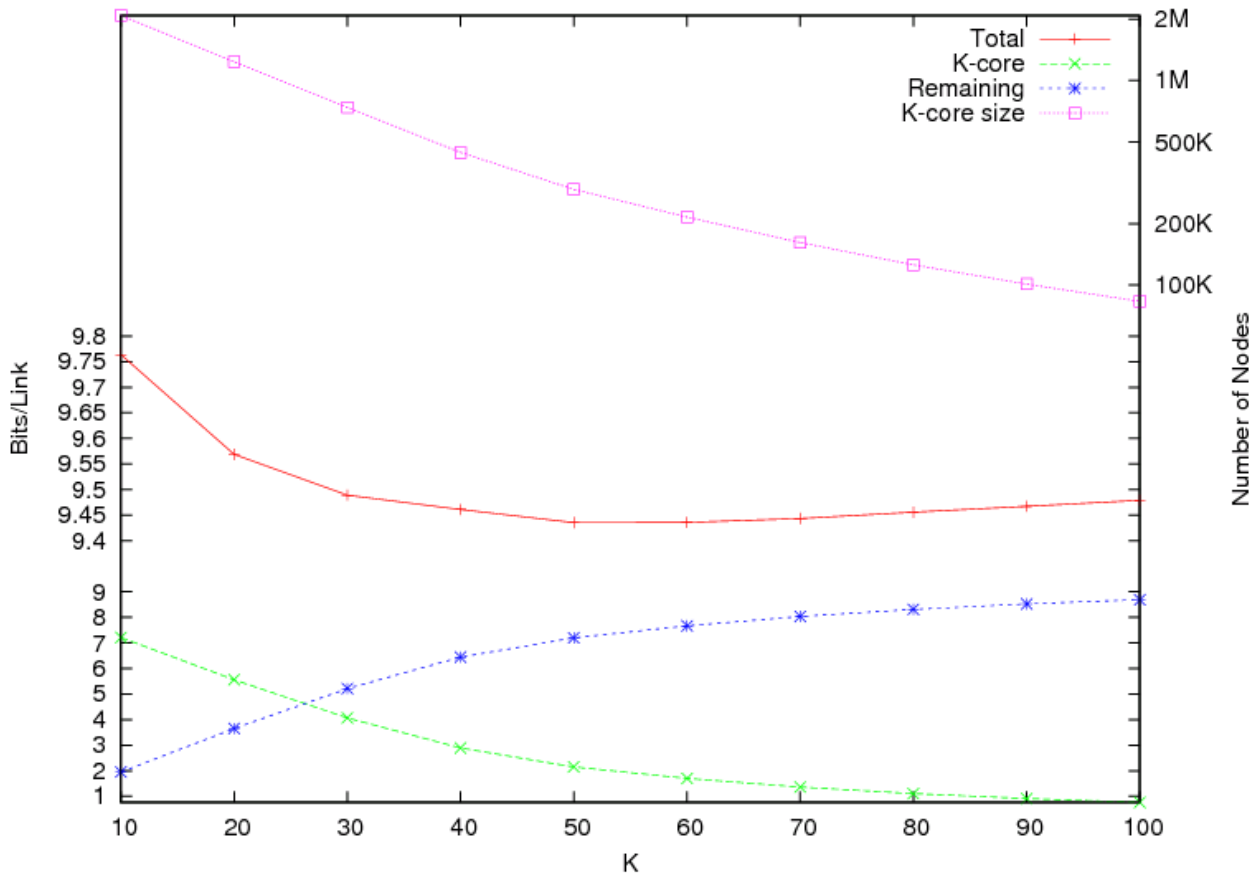
- **Preferential attachment model:** Rich get richer – a new node links to an existing node with probability proportional to its degree
- Shows that shingle ordering helps BV-style compressions in stylized graph models

Gap distribution



Shingle ordering produces smaller gaps

Who is the culprit



Low degree nodes are responsible for incompressibility

Incompressibility

Theorem. The following generative models all require $\Omega(\log n)$ bits per edge on average, even if the node labels are removed

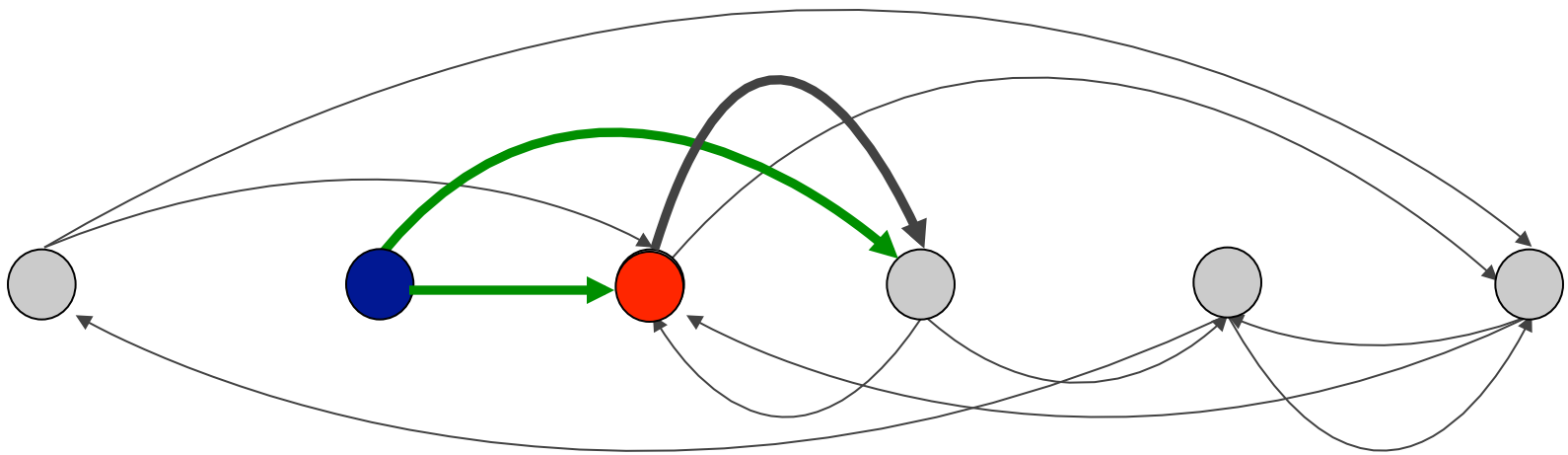
- the preferential attachment model
- the copying model
- the evolutionary ACL model Aiello, Chung, Lu FOCS 2001
- Kronecker multiplication model Leskovec et al PKDD 2005
- Model for navigability in social networks Kleinberg Nature 2000

- We remove labels since BV compresses unlabeled Web graphs to O(1) bits per edge
- Min-entropy argument: Find a subset of graphs
  - not too large: to avoid graphs that are "easy"
  - not too small: should still contain interesting graphs about which we can show incompressibility

A compressible graph model

- Begin with a seed graph of nodes with out-degree k, arranged in a cycle
- Additional nodes arrive in sequence
- An arriving node is inserted at a random place in the cycle
  - It links to k-1 out-neighbors of its cycle successor

An example, k=2



Locality in the new model

- If a web designer wants to add a new web page to her web site
  - likely to take some existing web page on her website
  - modify it as needed (perturbing the set of its outlinks) to obtain the new page
  - adding a reference to the old web page
  - and publish the new web page on her website
- Since web pages are sorted by URL in our ordering, the old and the new page will be close!

Basic properties of the model

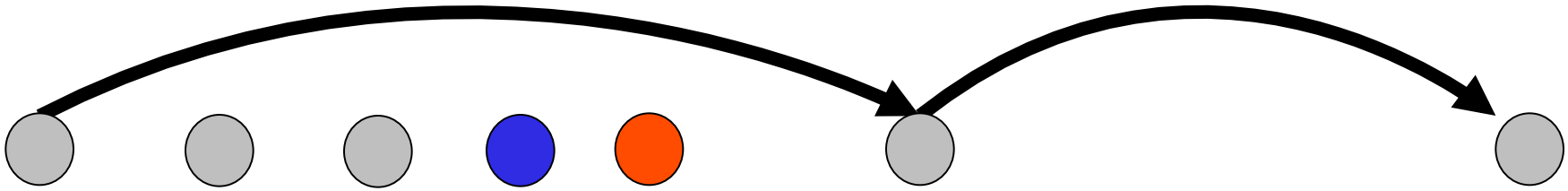
- Rich get richer: in the model, in-degrees converge to a power law with exponent -2-1/(k-1)
- High clustering coefficient
- Polynomially many bipartite cliques
- Logarithmic undirected diameter

- Compressible to O(1) bits per edge
- In fact, BV algorithm achieves O(1) bits per edge

Compressibility

- **Theorem.** The number of bits required by BV algorithm is $\sum_{l=1..\infty} Y_l (\log l)$, where Y_l is the number of edges of length l
- **Theorem.** In the model, edge lengths converge to a power law with exponent $-1-1/k$
- **Corollary.** The new model produces graphs compressible to O(1) bits per edge

Long gets longer



- Recall the process: pick a leader node uar and place new node to its immediate left
- The probability to become longer is proportional to the number of nodes "below" the edge, ie, its length
- Making this precise requires pinning down subtle combinatorial properties of the model

Summary

Temporal analysis of behavioral networks yields richer understanding of its dynamics

- **Degree distributions**
- **Diameter and density**
- **Components**
- **Compressibility**

Future directions

- General questions
  - Richer models
  - Surprising properties
  - Mathematical tools for analysis
- Specific questions
  - Conversational trees
  - Information spread, influence vs correlation
  - Can we compress social networks better?
  - Good algorithms for compression-friendly orderings

Thank you!

ravikumar@yahoo-inc.com