Information Networks and Semantics

Srinath Srinivasa
IIIT Bangalore
sri@iiitb.ac.in

Outline

- Information versus Material Networks
- Semantics in Information Networks
  - Knowledge Networks
    - Triadic Closure
    - Entrenchment
    - Affinity and Stability
  - Attention Networks
    - Dynamics of Attention
    - Interpreting Co-citations

Recent new additions to our vocabulary

- Telemedicine
- SMS/MMS
- E-learning
- Net Banking
- E-ticketing
- Blogging
- Tweeting
- GPL, Open-source
- Creative Commons
- ...

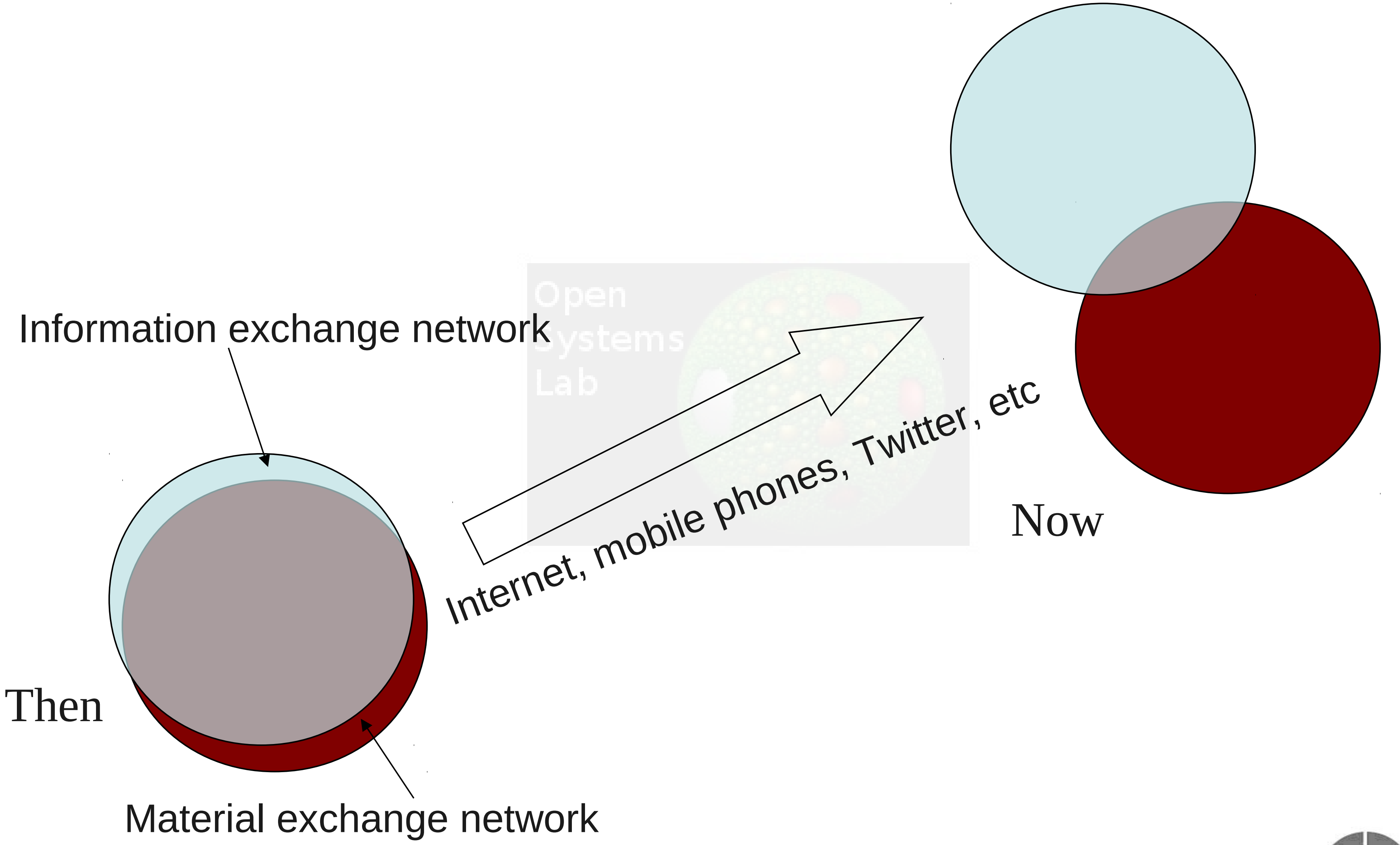
- Phishing
- Hacking
- Virus / Spyware / Adware / Malware
- Cyber squatting
- Cyber stalking
- Identity theft
- Piracy
- ...

The "Information Age"

- Comprehensive change brought by information and communication technologies (ICT)
- *Qualitative* changes affecting the underlying mental model or the "paradigm"
- Changes affecting the way we lead day-to-day lives (not just businesses)

- At the core:
  - Separation of information transactions from material transactions
  - Historically information transactions were "piggy backed" on material carriers (Ex: postal mail)
  - Information transactions over material carriers: Conduction, convection and radiation metaphors

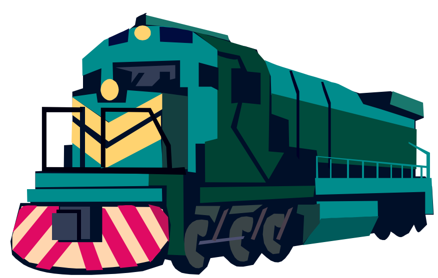
The “Information Age”

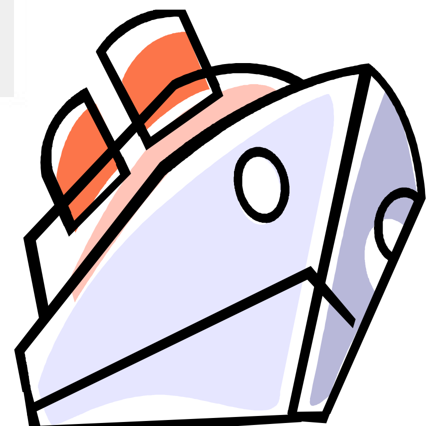


Material exchange

- Constrained by the laws of physics
- Conserved (zero-sum) transactions
- High cost of replication
- High cost of transportation

The Result:





Material exchange

- Characteristics of efficient material exchange
  - Standardization
  - Mass-production

Information exchange

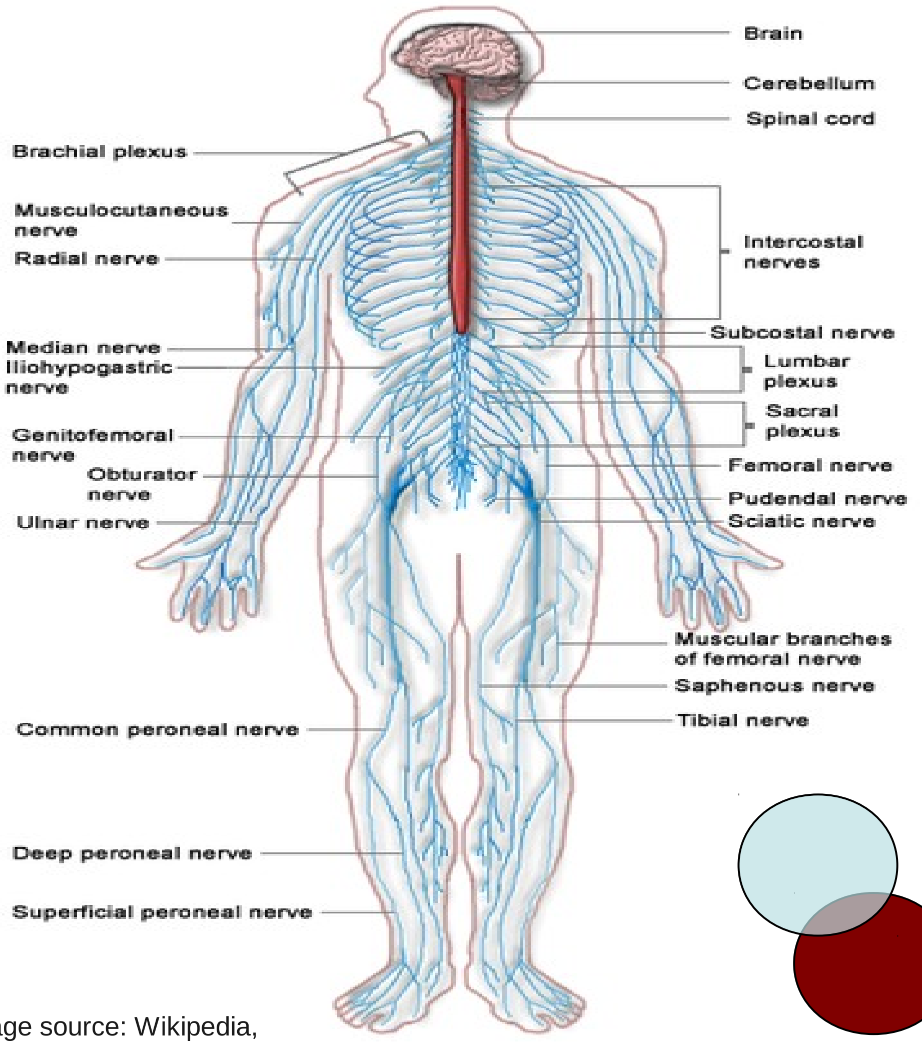
- Intangible artifacts
- Non-conserved transactions
- Very low cost of replication
- Very low cost of transport

The Result (until now, at least):

Information versus material

- Information exchanges are copy operations while material exchanges are move operations
  - If someone steals my car, I will come to know of it very soon; but if someone is reading my network traffic, I may never know
- Material affects *what you have*, while information affects *what you are*
  - “Wisdom” from old movies: The only way to handle someone who “knows too much” is, well…

Information networks in nature



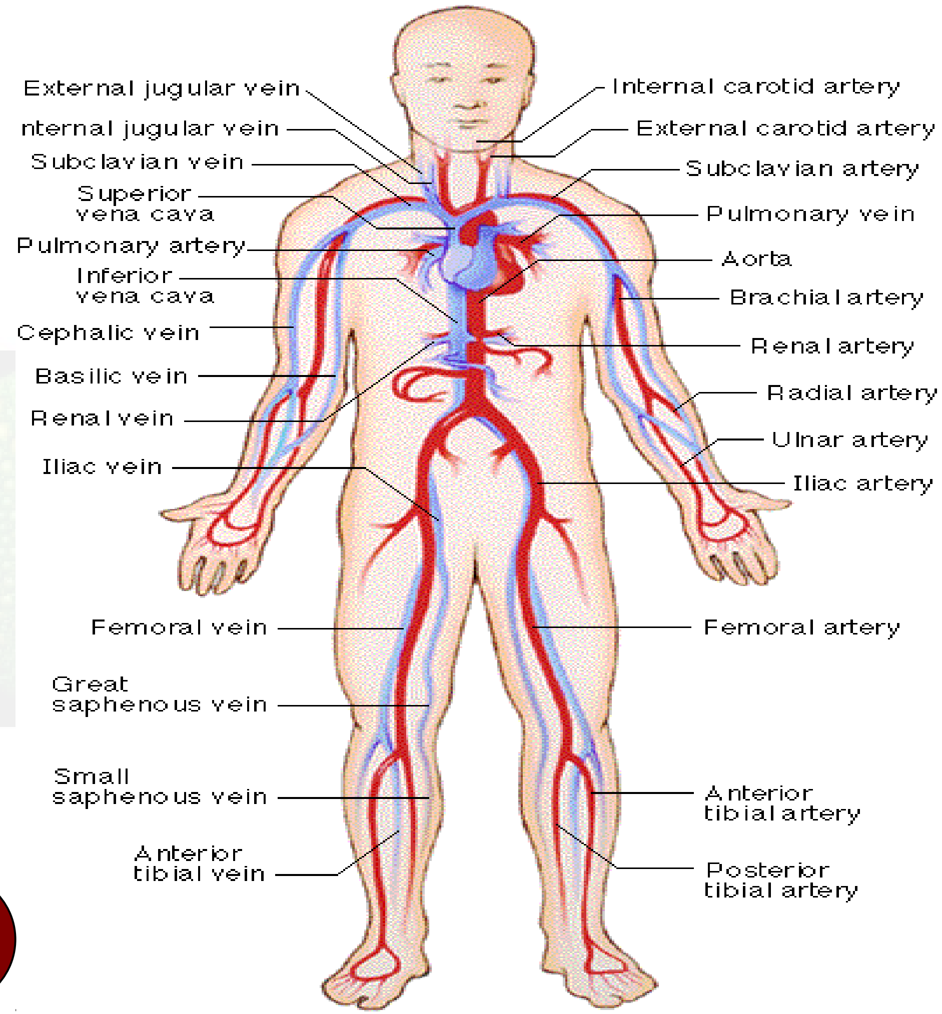
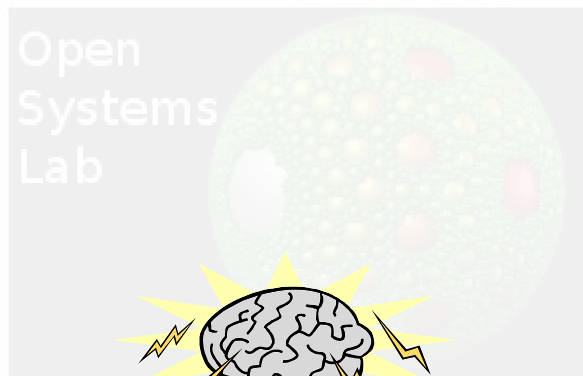


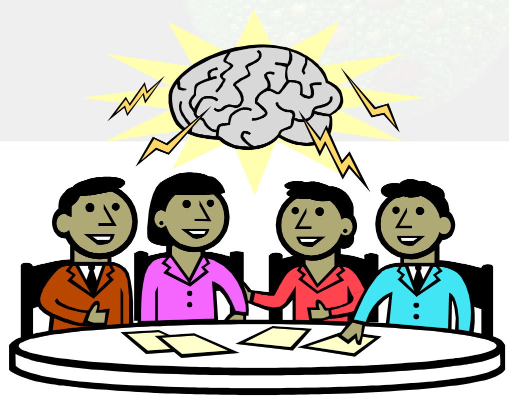
Image source: Wikipedia, HSTA

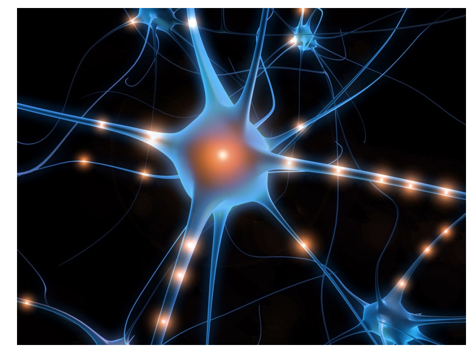
Note the separation between material and information logistics.
Nature is already in the information age!

Networks compared..

- Material networks
  - Exchange of resources (and in turn, energy)
  - Waste disposal
- Information networks
  - Control
  - Semantics







Information Networks

and

Semantics

Networks versus Graphs

- Network
  - A system of agents and interconnections using which elements are exchanged between agents. Interconnections are based on perceived value by an agent and change over time

- Graph
  - An abstract representation of an instantaneous state of a Network, in the form of nodes and edges.

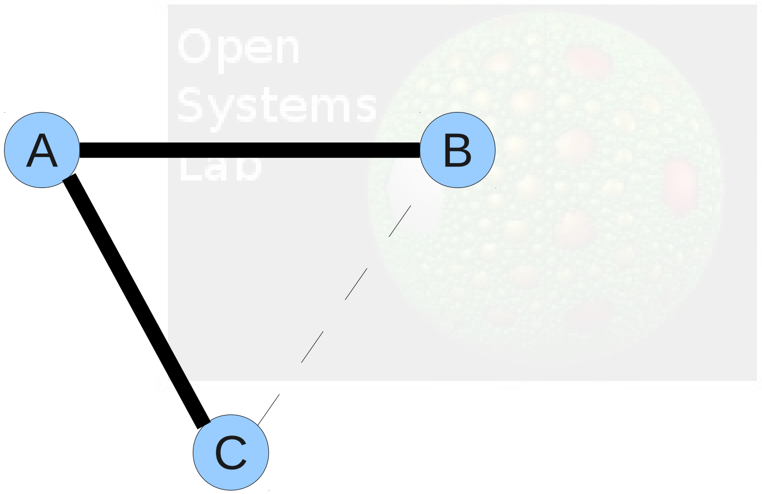
Knowledge Networks

Social Networks

- Interpretations of friendship connections:
  - Knowledge about each other (familiarity)
  - Time spent together
  - Resonance of opinions

Triadic Closure

- Regardless of the interpretation of a social link, social networks are hardly formed randomly
- A characteristic feature of social networks is the property of *Triadic Closure*



Informally: Two people who have a common friend are likely to become friends themselves. The more closer they are to their common friend, the more likely is it that they become friends themselves

Triadic Closure

Triadic closure is not a property of how people behave.. it is a *network property* of acquaintance relationships

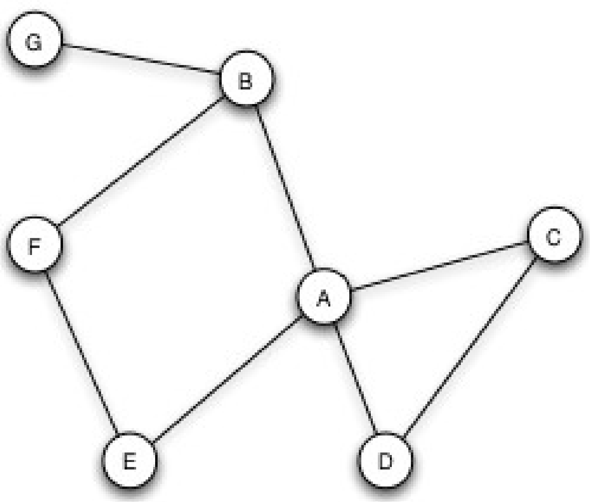
Consider a set of "issues" S and a set of actors A. Let S_a be the set of "opinions" actor a has about issues in S. Given this, the strength of associations is defined as:

$$w_{ab} = \frac{|S_a \cap S_b|}{|S|}$$

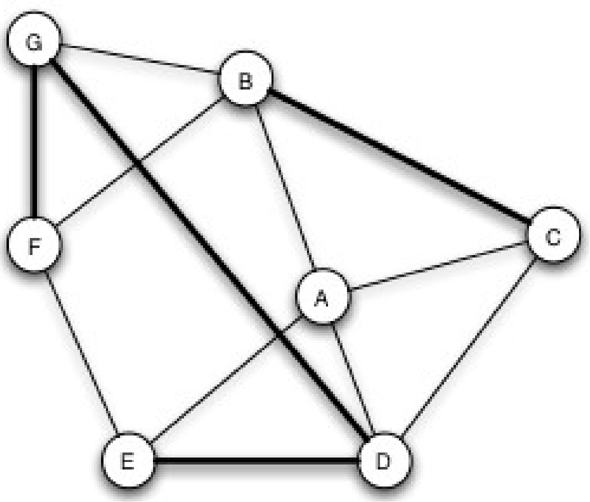
Given any two associations with strengths w_{ab} and w_{ac}, Triadic closure is imminent (pigeon-hole principle) when w_{ab} and w_{ac} is more than 0.5.

Entrenched Networks

Triadic closure property creates an effect of "entrenchment" in acquaintance networks



(a) *Before new edges form.*



(b) *After new edges form.*

Probability of learning something new diminishes with more entrenchment

Image Source: [Easley and Kleinberg '10]

Entrenched Networks

Formally

Given a graph G = (V,E) and a node v ∈ V, with its neighbourhood denoted by N(v), the entrenchment factor of the node is defined as:

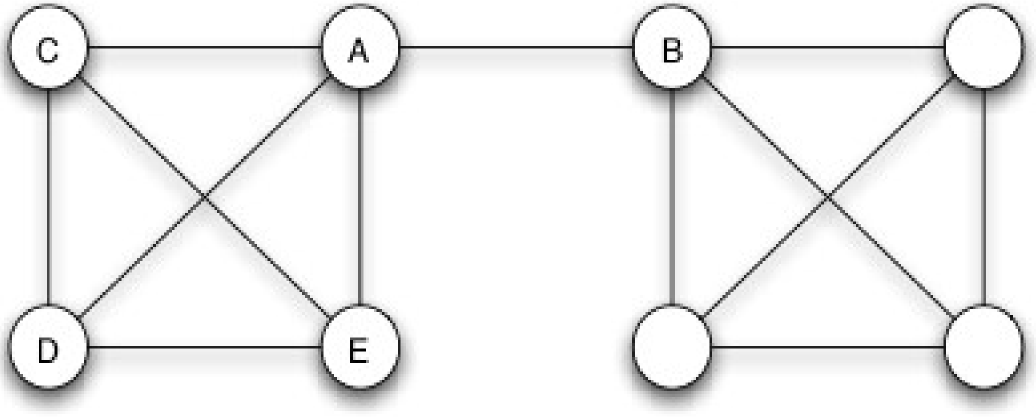
$$\mu_v = \frac{1 + |\{(u,w) \mid u,w \in N(v)\}|}{|N(v)|}$$

Entrenchment factor for the graph can be computed as the average of the entrenchment factors for all nodes

$$\mu_G = \frac{\sum\limits_{v \in V} \mu_v}{|V|}$$

The Strength of Weak Ties

Bridges



Bridges are links connecting two entrenched components of a network.

In social networks, bridges are necessarily “weak links” (acquaintances, rather than friendships) as strong ties would have resulted in triadic closure around the bridge..

Image Source: [Easley and Kleinberg '10]

The Strength of Weak Ties

- Bridges are primary sources of new knowledge and insight in entrenched networks
- A now famous study by [Granovetter '73] shows the importance of weak ties in social networks
- People make major decisions of their lives (career, marriage,...) based on connections obtained from weak ties, rather than strong ties
- Weak ties also a primary source of knowledge diffusion in online social networks (Retweets, Facebook shares, etc.)

The Strength of Weak Ties

Formally, bridges that are sources of knowledge are edges with high *betweenness centrality*:

$$c_B(e) = \sum_{\forall s,t \in V} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

...where σ_{st} is the number of shortest paths between nodes s and t,

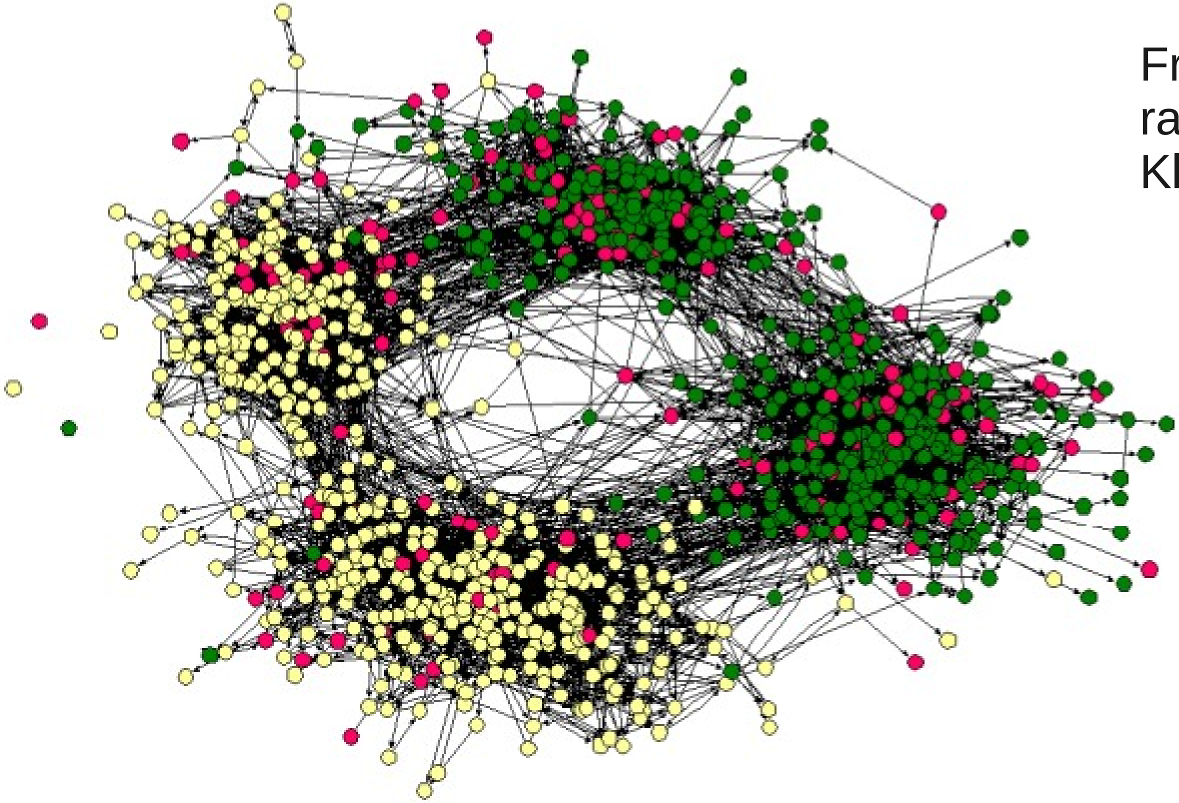
and $\sigma_{st}(e)$ is the number of shortest paths between nodes s and t that pass through e.

The Strength of Weak Ties

- Other properties of bridges:
  - Adding a bridge connecting two disconnected, entrenched components *increases* the diameter of the largest connected component (LCC)
  - Increased diameter has implications like:
    - Increased distortion in knowledge transfer
    - Reduced trust
    - Greater novelty (something new to learn) in individual interactions
  - Adding connections makes the world bigger before it becomes smaller

Homophily

Social networks based on knowledge, display a property of *Homophily* or attraction to others of the same type



Friendship networks based on race in a US school [Easley and Kleinberg '10]

Image Source: [Easley and Kleinberg '10]

Positive and Negative Relationships

- Affinity based on homophily can also have a *negative* component: disaffinity towards other kinds
- Examples: Disaffinity based on political ideologies, software platforms, religious beliefs, etc.
- Affinity and disaffinity can be modeled as positive and negative social relationships

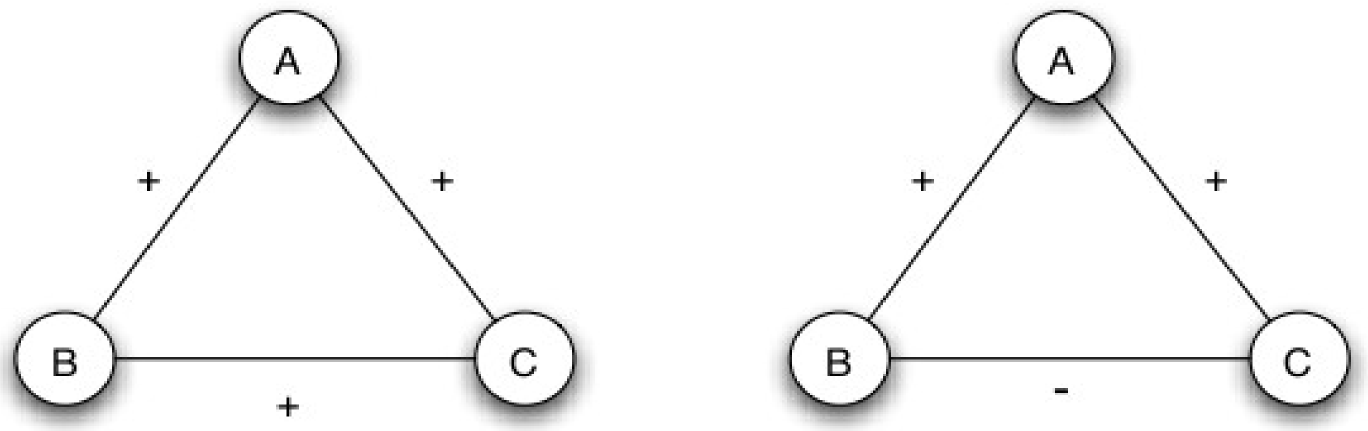
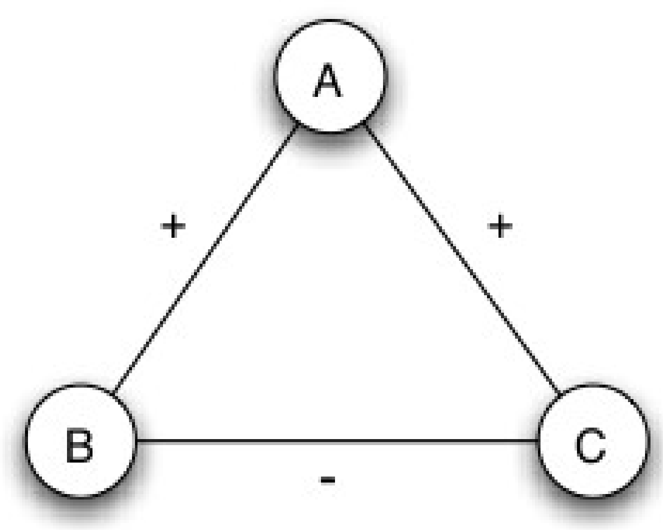


Image Source: [Easley and Kleinberg '10]

Stability

Disaffinity brings in the issue of network stability:

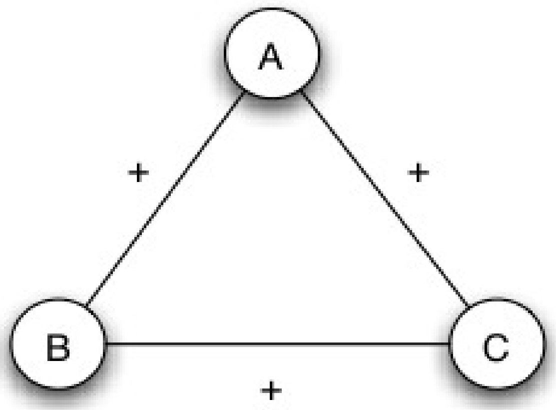


In the example network, triadic closure brings B and C together as the affinity between AB and AC increases.

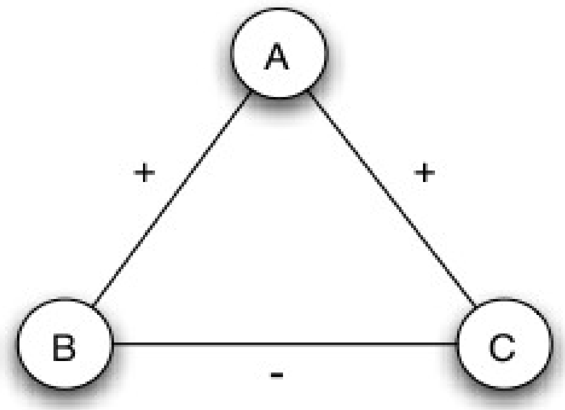
But since there is already disaffinity between B and C, this makes the network unstable..

Image Source: [Easley and Kleinberg '10]

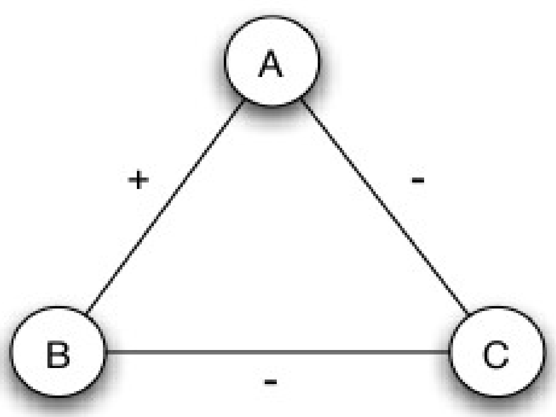
Structural Balance



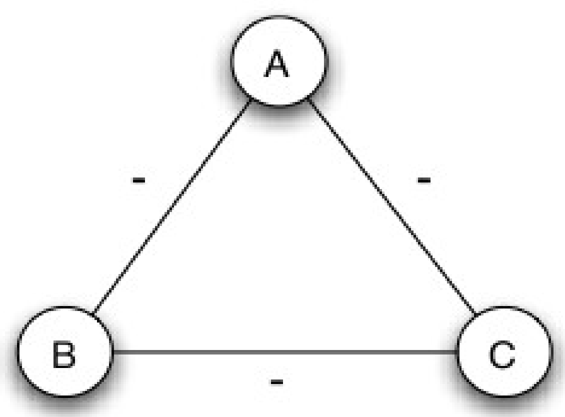
(a) *A, B, and C are mutual friends: balanced.*



(b) *A is friends with B and C, but they don't get along with each other: not balanced.*



(c) *A and B are friends with C as a mutual enemy: balanced.*

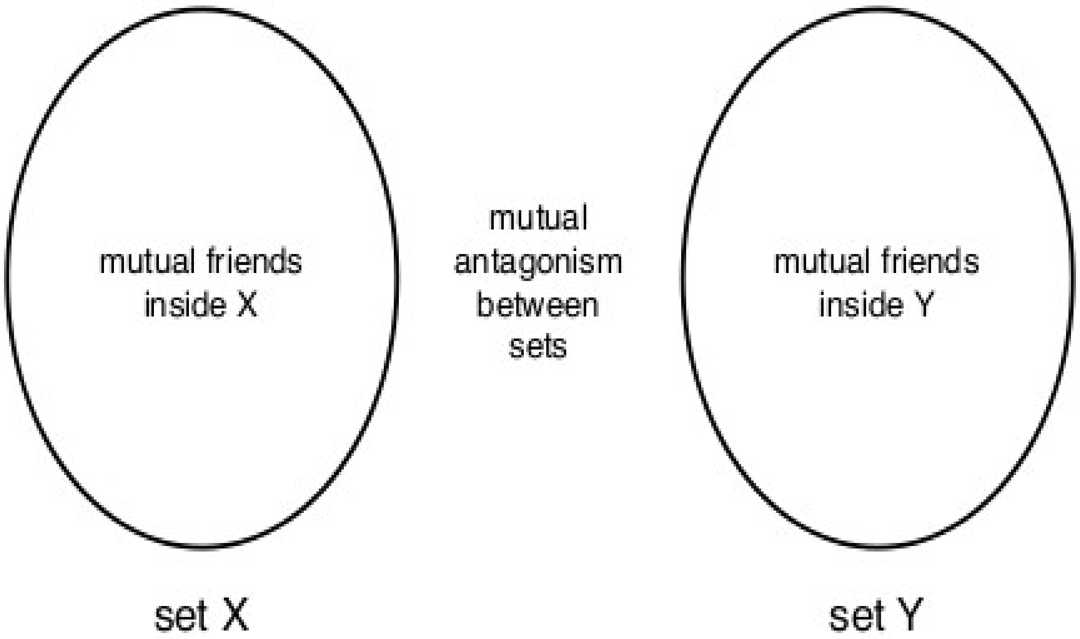


(d) *A, B, and C are mutual enemies: not balanced.*

Weakly balanced

Image Source: [Easley and Kleinberg '10]

Cartwright-Harary Theorem



A complete graph is (strongly) balanced iff either all nodes have affinity to all others, or the graph can be divided into two groups, each of which have affinity among themselves and disaffinity with all members of the other group

Proof beyond the scope of this talk..

Image Source: [Easley and Kleinberg '10]

Attention Networks

The Economics of “Attention”

[Goldhaber 1997]

- The commodity of scarcity in an information-rich world: *Attention*
- Takes on different forms in media studies: Eyeballs, Immersiveness, PageRank, etc.

- Attention: The process of concentrating on one piece of information from the environment while ignoring all others
- Opposite of distraction.

Why does attention matter?

Image Source: Wikipedia

Why does Attention Matter?

- *Did you ever imagine you'd be thinking about Himalayas / Tibet when you decided to attend this talk?*
- Attention controls what gets loaded into conscious memory
- Attention is a conserved entity unlike information, and hence is scarce
- Extended attention form associations, whether or not they represent real associations found in nature

Modalities of Attention Transfer Online

- Hyperlinks
  - User clicking on a hyperlink transfers attention from host page to target page

- Twitter followers / subscriptions
  - List of followers of a person indicates people who wish to pay attention to a person's tweets

Other Interpretations

- Hyperlinks
  - Endorsement
  - Recommendation (advocacy)
  - Navigation
- Subscriptions
  - Endorsement
  - Recommendation
  - Acquaintance

- Links versus Citations
  - Links are *navigational* constructs facilitating attention transfer (typically nepotistic citations)
  - Citations are attention transfer constructs based on the *content / topic* in question
  - Most subscription / follower networks are links rather than citations

Modeling Attention Dynamics

PageRank

The PageRank or *Eigenvector centrality* is the most common way of determining aggregate attention that a page receives.

PageRank of a node v in a directed graph, is the probability that any random walk on the graph will feature v in the walk

$$\rho(v) = \frac{1}{|N_I(v)|} \sum_{u_i \in N_I(v)} \rho(u_i)$$

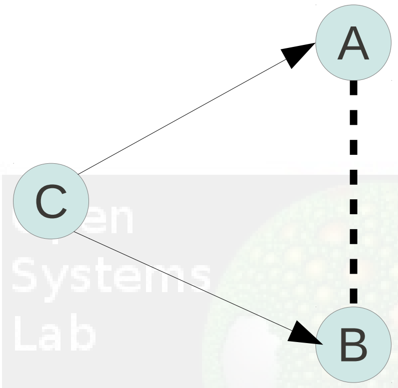
where

$$N_I(v) = \{(u, v) \mid (u, v) \in E\}$$

Modeling Attention Dynamics

- PageRank of a page represents the probability of a random surfer visiting the page in an infinitely long random walk
- Random surfer assumed to click on each hyperlink uniformly at random
- However, hyperlinks are not created equal – be they links or citations

Co-citations

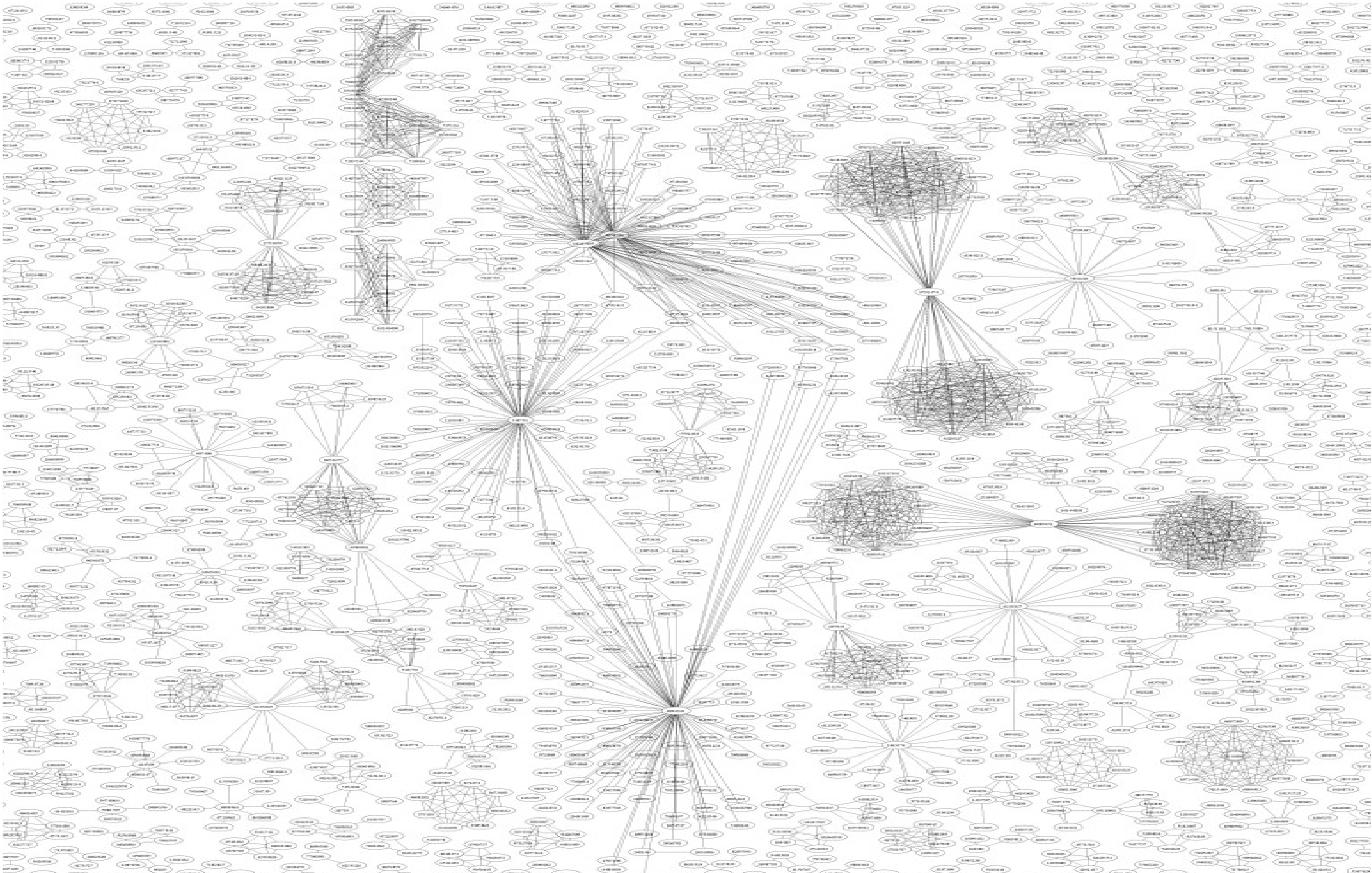


Two pages (people) A and B are said to be co-cited (co-followed) if there are one or more other pages (people) that cite (follow) both A and B.

A co-citation graph C(G) of a given hyperlink graph G = (V,E) is a weighted, undirected graph showing co-citations across pairs of pages and their counts.

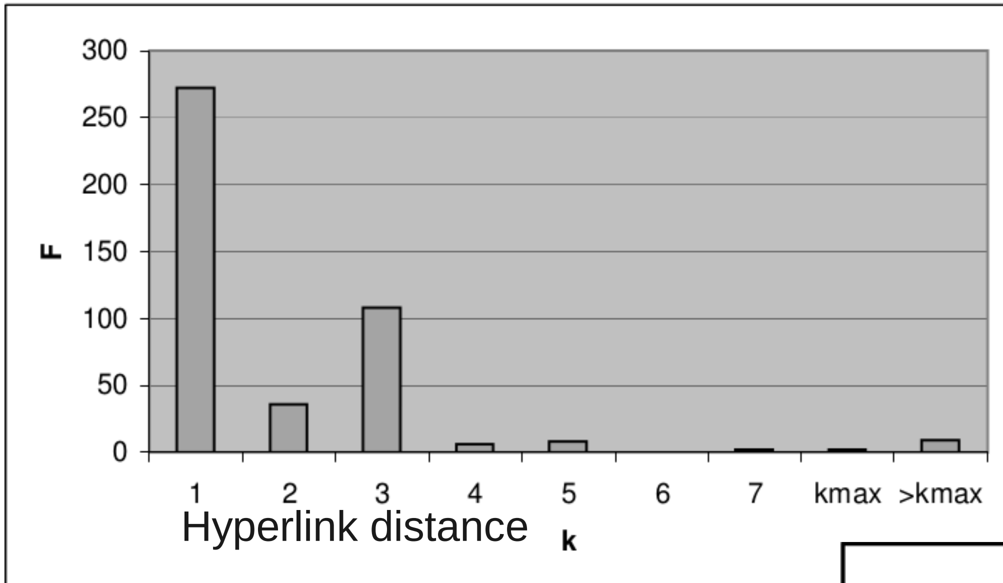
$$C(G) = (V, E, w)$$

Co-citation graph from a web crawl sample



Co-citations in the Web and Wikipedia

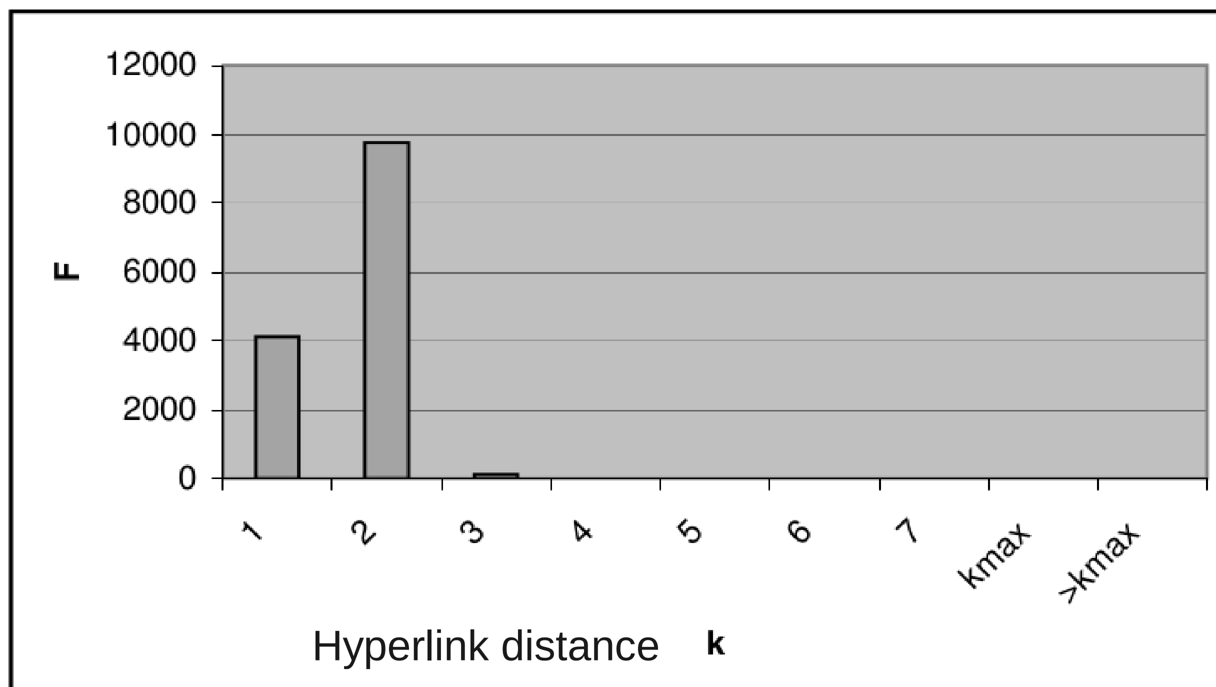
co-citations



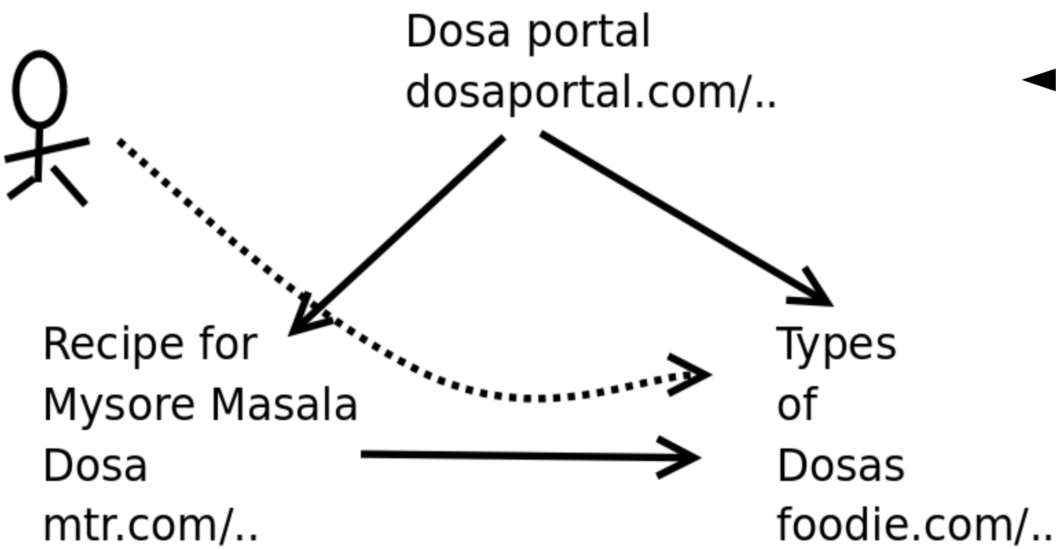
On the Web

Modal hyperlink distance between pairs of highly co-cited pages [Reddy et al '06]

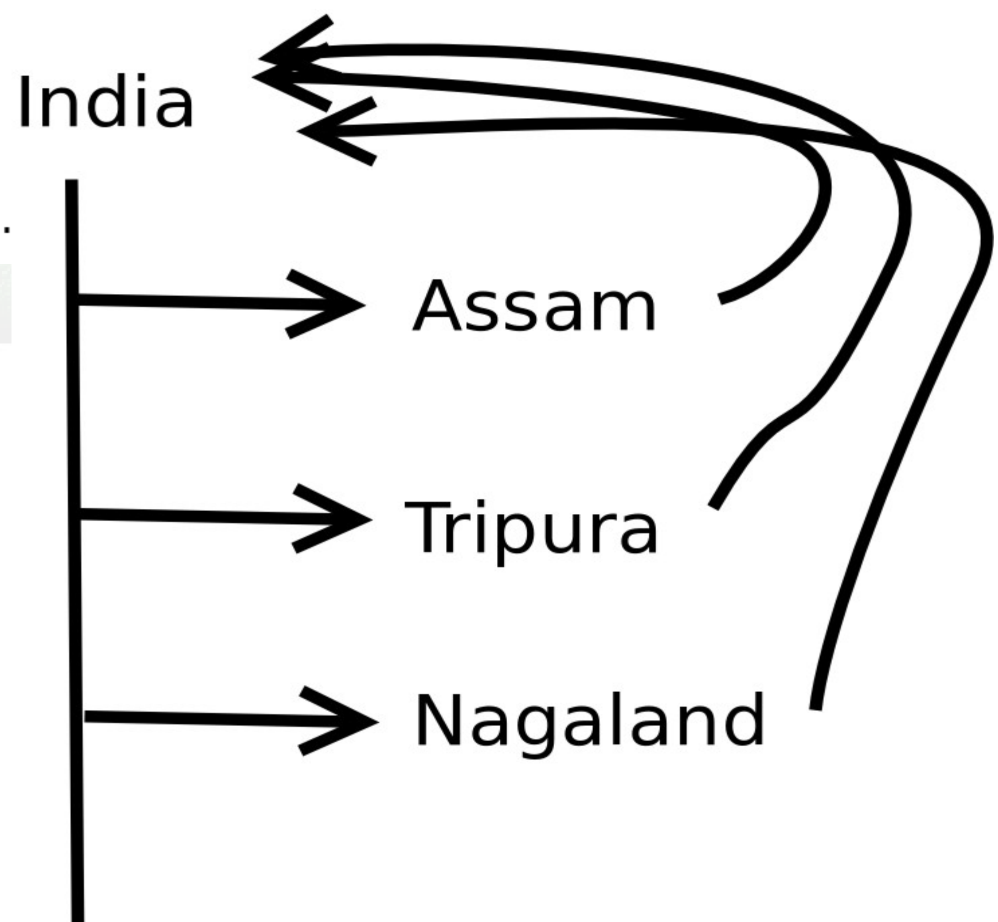
On Wikipedia



Interpreting Co-citations



Co-citation on the Web:
Citation Endorsements



Co-citation on Wikipedia:
Topical Aggregation

Citation Endorsements

- A significant form of co-citations on the web are due to prior existence of a citation
- If a page A has several citations, but only a few of the cited pages are also co-cited with A, they can be viewed as *third-party endorsements* of citations
- Formally:

$$w(A \rightarrow B) = \frac{|Coci(A \rightarrow B)|}{\sum_{\forall B, A \in N_I(B)} |Coci(A \rightarrow B)|}$$

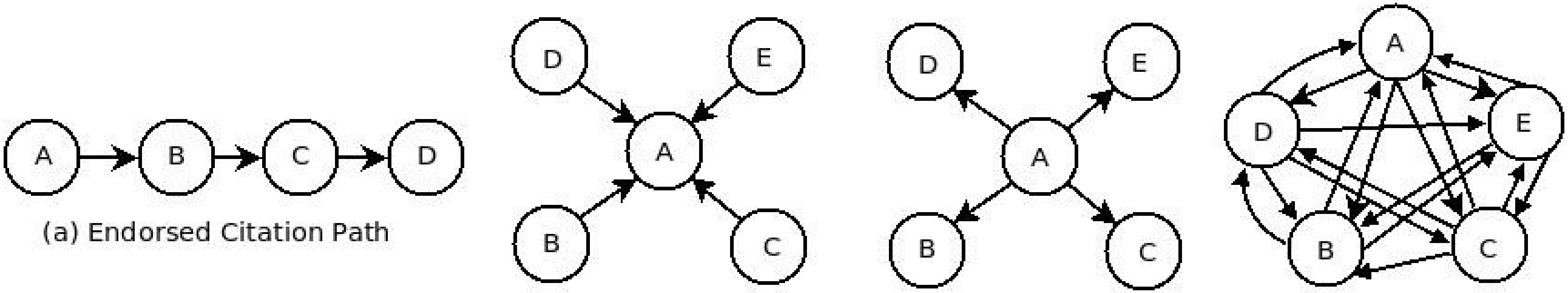
where Coci() represents the set of all co-citations of a hyperlink..

Attention Backbone

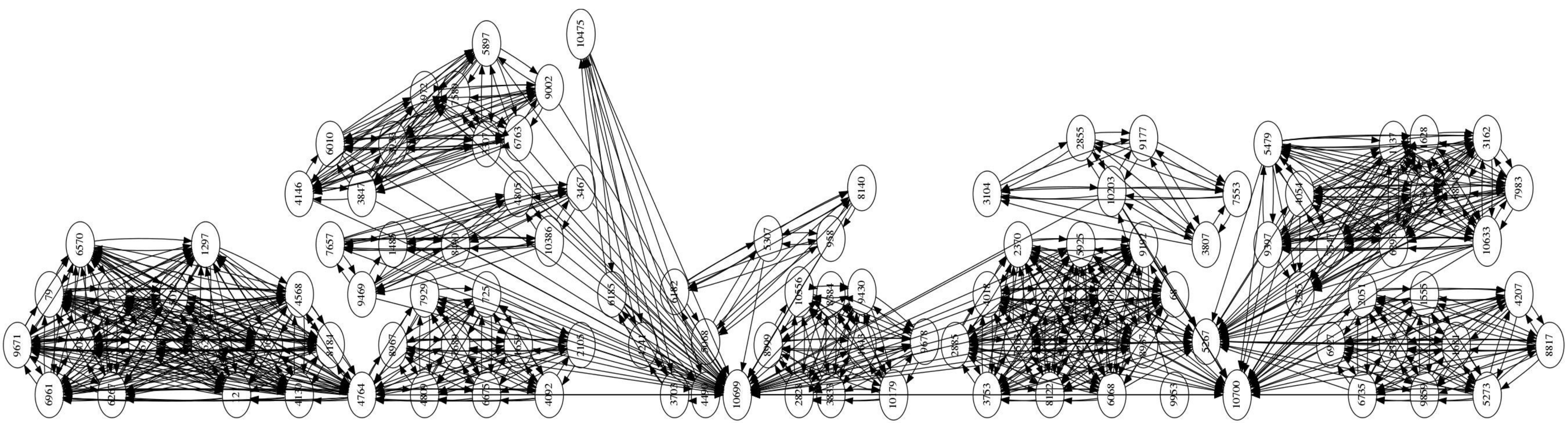
- Citations represent flow of attention in a hyperlinked repository
- Endorsed citations distinguish between “highways” and “bylanes” in the attention flow network
- The hyperlink graph formed by only highly endorsed hyperlinks called the Endorsed Citation Graph (ECG), forms the “Attention backbone” of the hyperlink repository
- ECG for web crawl shown to be made of disconnected components with a skewed distribution of component sizes

Web Crawl ECG

Structural Motifs



(a) Endorsed Citation Path (b) Authority-Spoke Pattern (c) Hub-Spoke Pattern (d) Dense Cluster



ERank

- ERank is an authority score of a page within an ECG component
- Depicts reachability of the page within the component
- ERank scores in a component shown to be *uncorrelated* to the PageRank scores of pages of that component
- ERank provides indication of interestingness of page to a topical surfer

Conclusions

- Information Networks: Control and Semantics
- Units of information network dynamics:
  - Attention
  - Affinity
  - Knowledge
  - ...?
- Several open problems to be explored..

Thank You!

References

David Easley, Jon Kleinberg. Networks, Crowds and Markets: Reasoning about a Highly Connected World. Cambridge University Press, 2010.

Michael H. Goldhaber. The Attention Economy and the Net. First Monday, Volume 2, Number 4 - 7 April 1997.

Mark S. Granovetter. The Strength of Weak Ties. American Journal of Sociology, Volume 78, Issue 6 (May, 1973), 1360-1380.

Mandar R. Mutalikdesai, Srinath Srinivasa. Co-citations as Citation Endorsements and Co-links as Link Endorsements. Journal of Information Science, volume 36, issue 3, pages 383–400, 2010.

Siddhartha Reddy K, Srinath Srinivasa, Mandar R. Mutalikdesai. Measures of "Ignorance" on the Web. Proc. of COMAD 2006, New Delhi, December 2006.