CSE 258 – Lecture 12

Web Mining and Recommender Systems

Social networks

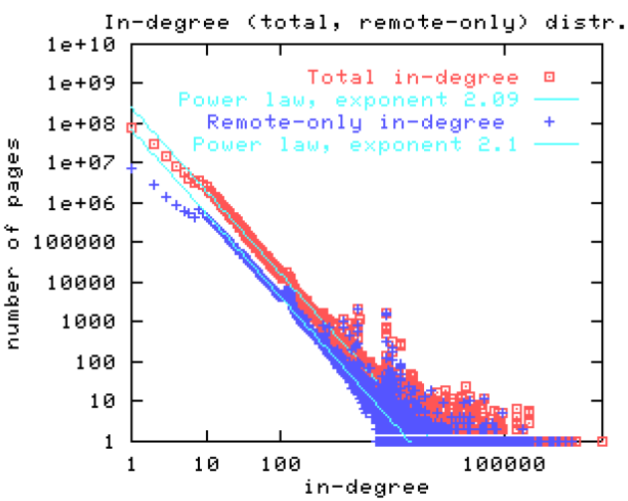
Social networks

We've already seen networks (a little bit) in week 3

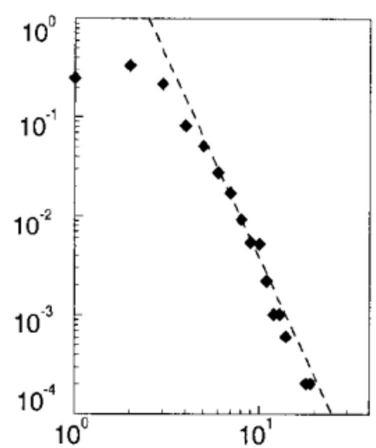
- i.e., we've studied inference problems defined on graphs, and dimensionality reduction/community detection on graphs
- **Q:** what do social & information networks **look like?**
- **Q:** how can we build better **models** that are tailored to the properties of social networks?

Social networks

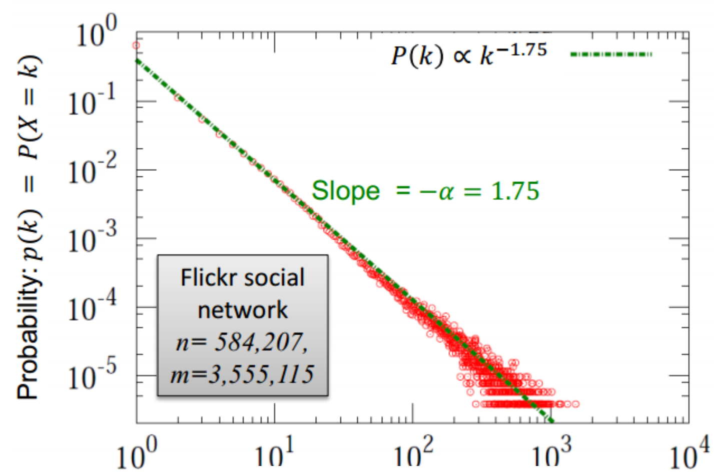
- Social and information networks often follow **power laws**, meaning that a few nodes have **many** of the edges, and many nodes have **a few** edges



e.g. web graph (Broder et al.)



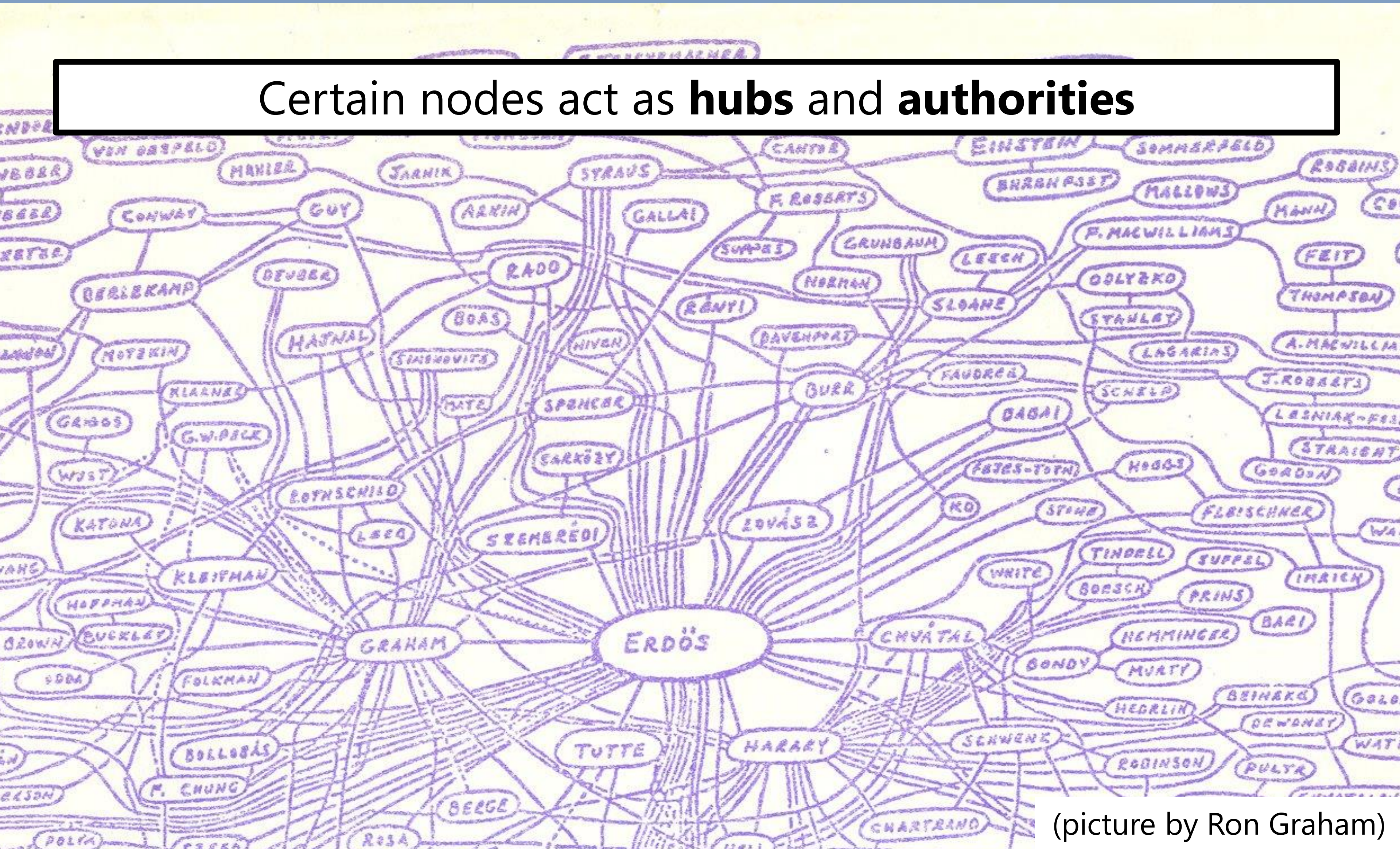
e.g. power grid (Barabasi-Albert)



e.g. Flickr (Leskovec)

Social networks

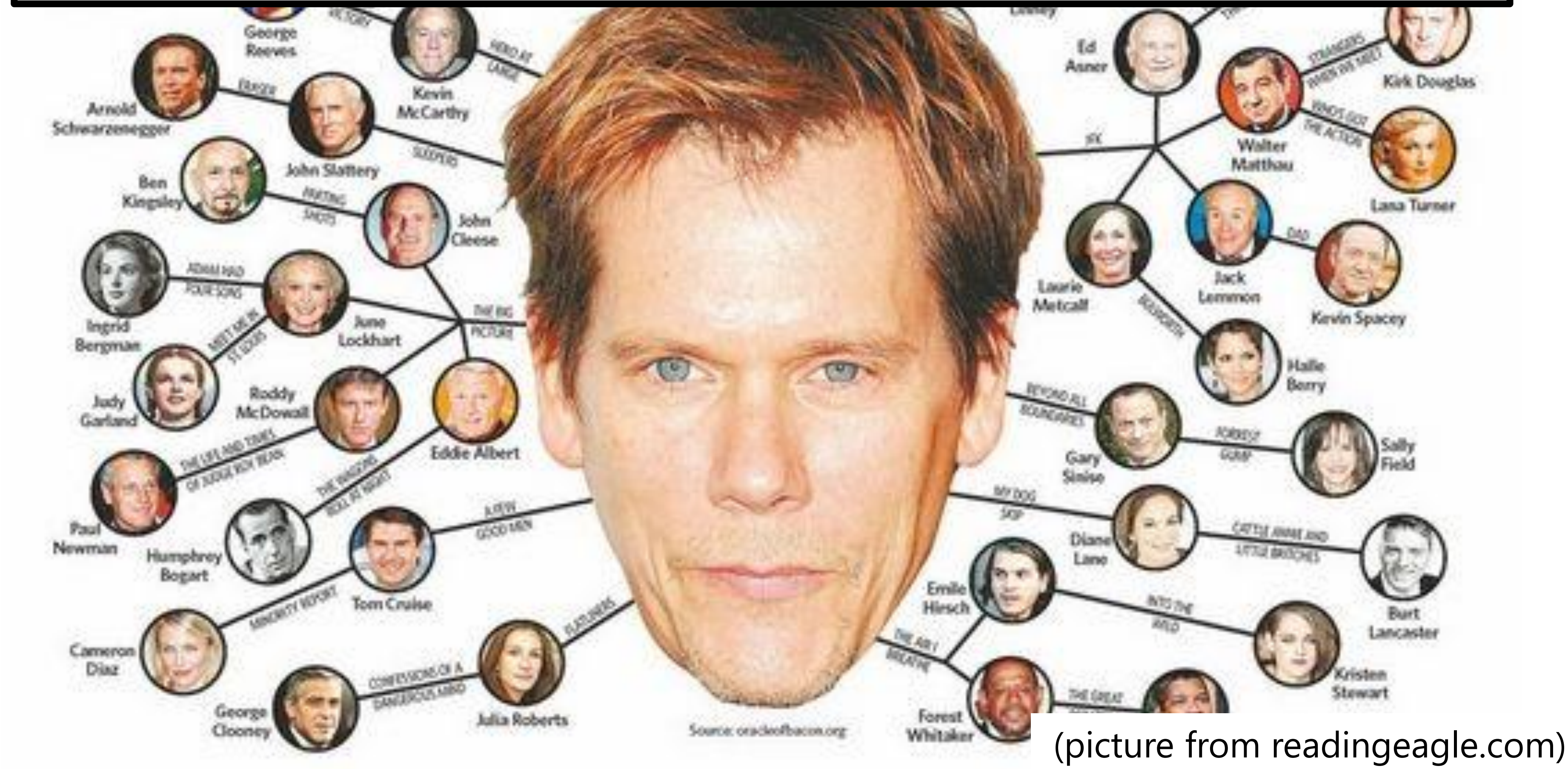
Certain nodes act as **hubs** and **authorities**



(picture by Ron Graham)

Social networks

Social networks are **small worlds:** (almost) any node can reach any other node by following only a few hops



(picture from readingeagle.com)

Social networks

How can we **characterize, model,** and **reason about** the structure of social networks?

1. Models of network structure
2. Power-laws and scale-free networks, “rich-get-richer” phenomena
3. Triadic closure and “the strength of weak ties”
4. Small-world phenomena
5. Hubs & Authorities; PageRank (maybe)

Social networks

How can we **characterize, model,** and **reason about** the structure of social networks?

- This topic is not discussed in Bishop, and is covered only a little bit in Charle's Elkan's notes (Chapter 14)
- For this lecture I more closely followed Kleinberg & Easley's book "Networks, Crowds, and Markets"
- (A pre-publication draft of) this book is available **(for free!)** on the author's webpage:

http://www.cs.cornell.edu/home/kleinber/networks-book/

Social networks

See also: entire classes devoted to this topic (maybe I'll teach one some day…)

NETS 112 "Networked Life"
(Michael Kearns @ UPenn)

cs224w "Social & Information Network Analysis"
(Jure Leskovec @ Stanford)

SAY BIG DATA ONE MORE TIME

ONLY WENT TO FIRST CLASS, MIDTERM, AND FINAL STILL GOT AN A

NETWORKED LIFE
Networked and Social Systems Engineering (NETS) 112
Fall 2014
Tuesdays and Thursdays 10:30-12, Berger Auditorium, Skirkanich Hall
Prof. Michael Kearns

Jump to the course schedule.

COURSE DESCRIPTION

- *What science underlies companies like Facebook, Google, and Twitter?*
- *What are the economics of email spam?*
- *Why do some social networking services take off, and others die?*
- *What do game theory and the Paris subway have to do with Internet routing?*
- *How does Google find what you're looking for... and exactly how do they make money doing so?*
- *What structural properties might we expect any social network to have?*
- *How might a social network influence election outcomes?*
- *What problems can be solved by crowdsourcing?*
- *How does your position in a social network (dis)advantage you?*

Networked Life looks at how our world is connected -- socially, strategically and technologically -- and why it matters.

The answers to the questions above are related. They have been the subject of a fascinating intersection of disciplines, including computer science, physics, psychology, sociology, mathematics, economics and finance. Researchers from these areas all strive to quantify and explain the growing complexity and connectivity of the world around us, and they have begun to develop a rich new science along the way.

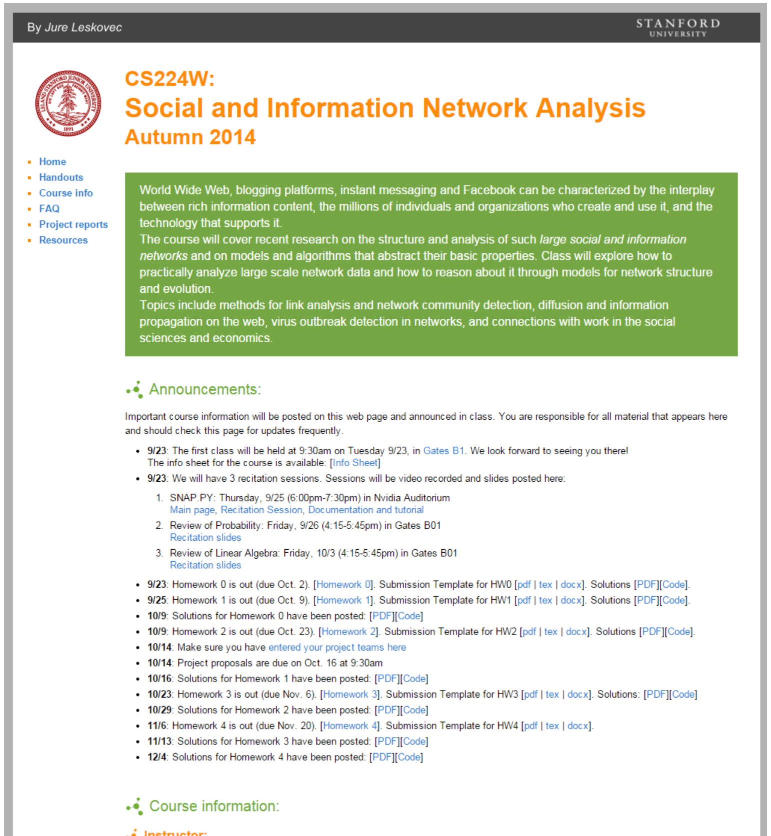
Networked Life will explore recent scientific efforts to explain social, economic and technological structures -- and the way these structures interact -- on many different scales, from the behavior of individuals or small groups to that of complex networks such as the Internet and the global economy.

This course covers computer science topics and other material that is mathematical, but all material will be presented in a way that is accessible to an educated audience with or without a strong technical background. *The course is open to all majors and all levels, and is taught accordingly.* There will be ample opportunities for those of a quantitative bent to dig deeper into the topics we examine. The majority of the course is grounded in scientific and mathematical findings of the past two decades or less (often much less).

Fall 2014 is the eleventh offering of *Networked Life.* You can get a detailed sense for the course by visiting the extensive course web pages from past years: [Fall 2013] [Fall 2012] [Fall 2011] [Spring 2010] [Spring 2009] [Spring 2008] [Spring 2007] [Spring 2006] [Spring 2005] [Spring 2004]
(Note: the Fall 2011 version used a different course management platform than the simple HTML site we'll be using this year, so it might be easiest to peruse the 2013 and pre-2011 sites to get a sense of how the course unfolds.)

There is also a greatly condensed version of this class offered to the general public as part of the online education platform Coursera. All Penn students should create a (free) Coursera account, and sign up for the session of Networked Life there that begins on Monday, September 1, 2014. See the course schedule for information on how we will make use of the online material and how to sign up.

Networked Life is the flagship course for Penn Engineering's recently launched Networked and Social Systems Engineering (NETS) program. Throughout the course we will foreshadow material that is covered in greater depth in later NETS program courses.



By *Jure Leskovec* STANFORD UNIVERSITY

CS224W:
Social and Information Network Analysis
Autumn 2014

- Home
- Handouts
- Course info
- FAQ
- Project reports
- Resources

World Wide Web, blogging platforms, instant messaging and Facebook can be characterized by the interplay between rich information content, the millions of individuals and organizations who create and use it, and the technology that supports it.
The course will cover recent research on the structure and analysis of such *large social and information networks* and on models and algorithms that abstract their basic properties. Class will explore how to practically analyze large scale network data and how to reason about it through models for network structure and evolution.
Topics include methods for link analysis and network community detection, diffusion and information propagation on the web, virus outbreak detection in networks, and connections with work in the social sciences and economics.

Announcements:

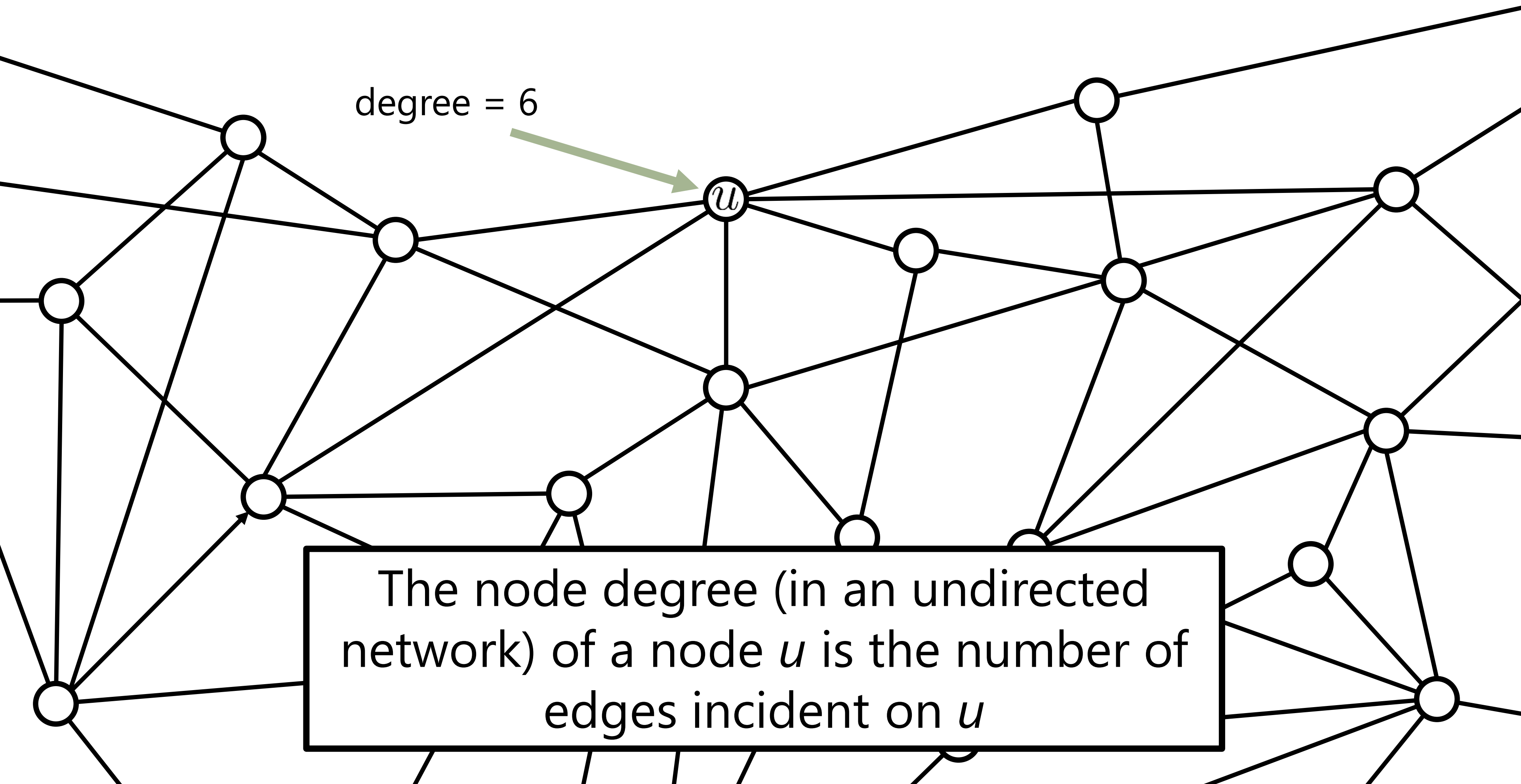
Important course information will be posted on this web page and announced in class. You are responsible for all material that appears here and should check this page for updates frequently.

- **9/23**: The first class will be held at 9:30am on Tuesday 9/23, in Gates B1. We look forward to seeing you there!
  The info sheet for the course is available: [Info Sheet]
- **9/23**: We will have 3 recitation sessions. Sessions will be video recorded and slides posted here:
  1. SNAP.PY: Thursday, 9/25 (6:00pm-7:30pm) in Nvidia Auditorium
     Main page, Recitation Session, Documentation and tutorial
  2. Review of Probability: Friday, 9/26 (4:15-5:45pm) in Gates B01
     Recitation slides
  3. Review of Linear Algebra: Friday, 10/3 (4:15-5:45pm) in Gates B01
     Recitation slides
- **9/23**: Homework 0 is out (due Oct. 2). [Homework 0]. Submission Template for HW0 [pdf | tex | docx]. Solutions [PDF][Code].
- **9/25**: Homework 1 is out (due Oct. 9). [Homework 1]. Submission Template for HW1 [pdf | tex | docx]. Solutions [PDF][Code].
- **10/9**: Solutions for Homework 0 have been posted: [PDF][Code]
- **10/9**: Homework 2 is out (due Oct. 23). [Homework 2]. Submission Template for HW2 [pdf | tex | docx]. Solutions [PDF][Code].
- **10/14**: Make sure you have entered your project teams here
- **10/14**: Project proposals are due on Oct. 16 at 9:30am
- **10/16**: Solutions for Homework 1 have been posted: [PDF][Code]
- **10/23**: Homework 3 is out (due Nov. 6). [Homework 3]. Submission Template for HW3 [pdf | tex | docx]. Solutions: [PDF][Code]
- **10/29**: Solutions for Homework 2 have been posted: [PDF][Code]
- **11/6**: Homework 4 is out (due Nov. 20). [Homework 4]. Submission Template for HW4 [pdf | tex | docx].
- **11/13**: Solutions for Homework 3 have been posted: [PDF][Code]
- **12/4**: Solutions for Homework 4 have been posted: [PDF][Code]

Course information:

Instructor:

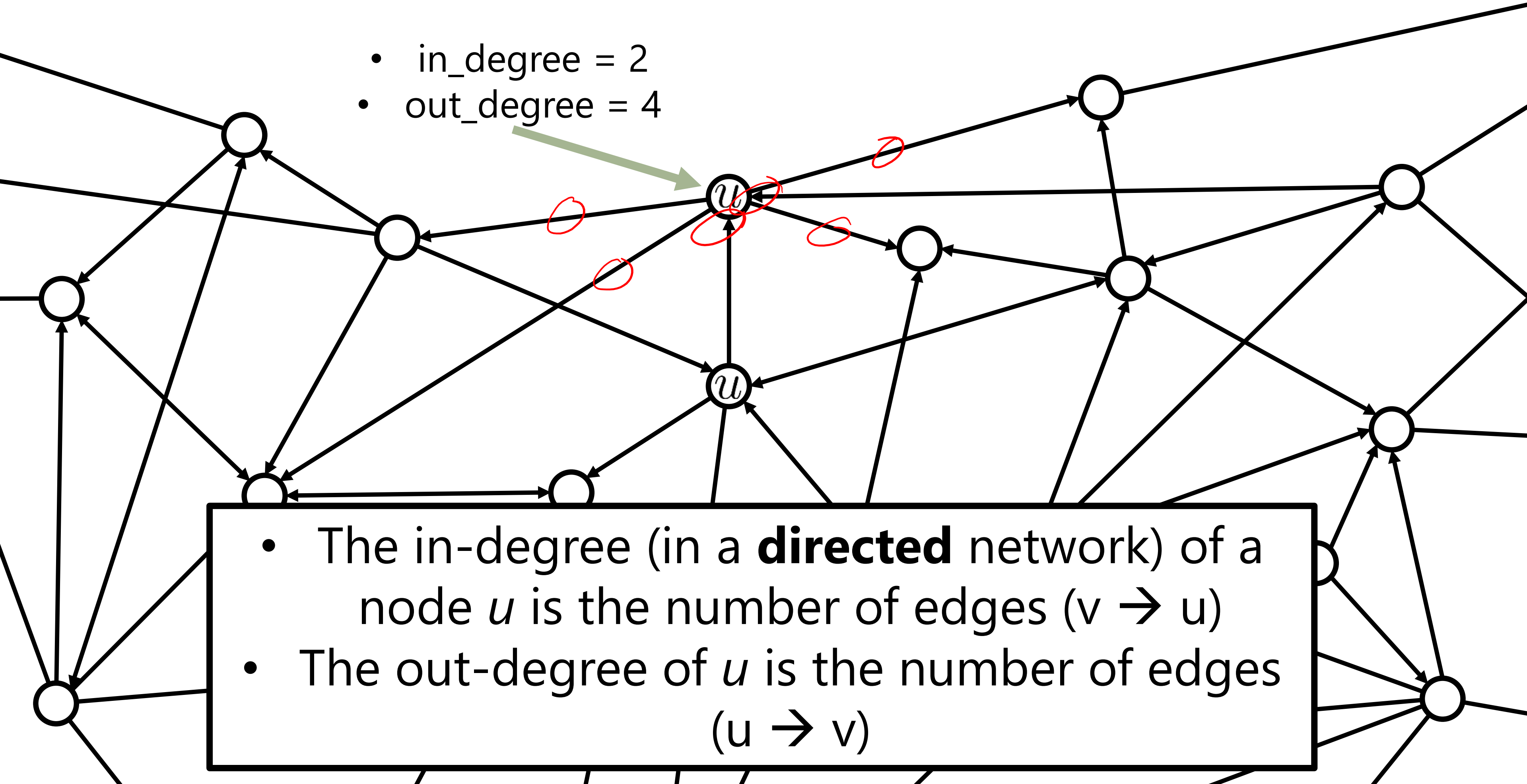
1. Node degree



The node degree (in an undirected network) of a node *u* is the number of edges incident on *u*

Definitions

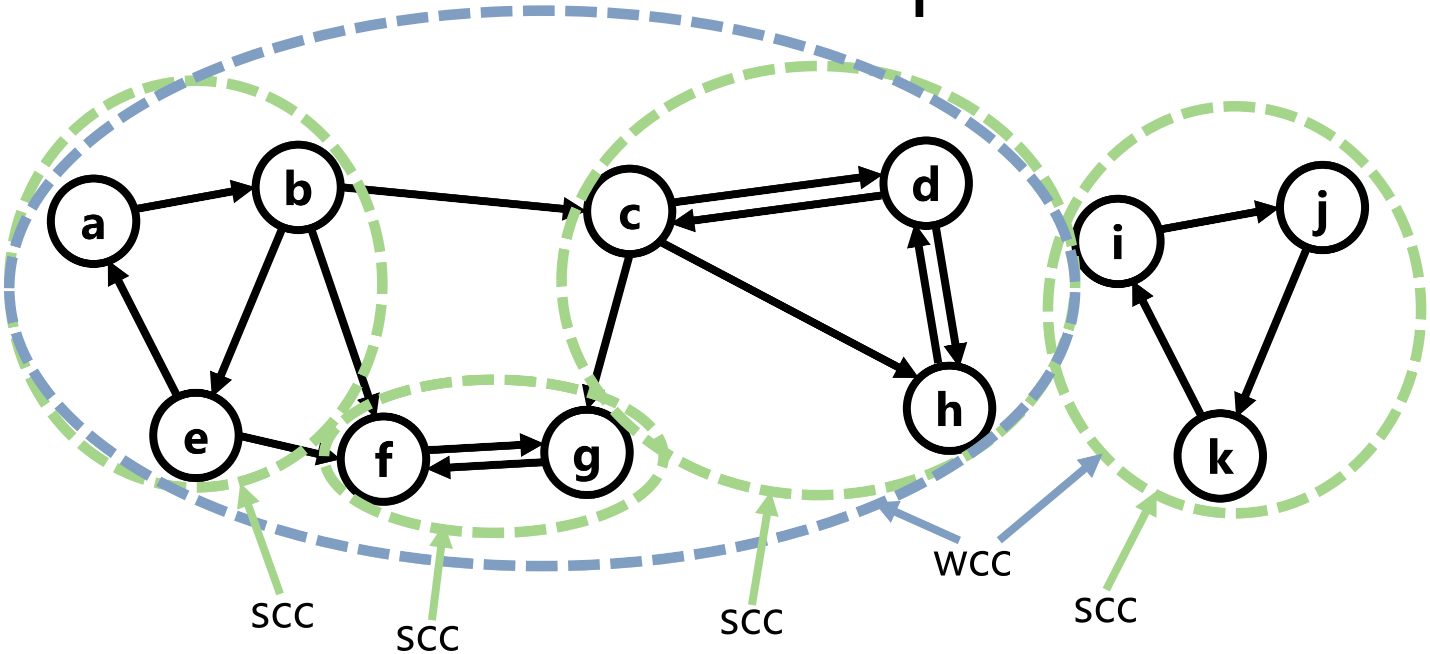
1. Node degree



- The in-degree (in a **directed** network) of a node *u* is the number of edges (v → u)
- The out-degree of *u* is the number of edges (u → v)

Definitions

2. Connected components



- If there is a path from (a→b) **and** from (b→a) then they belong to the same **strongly connected component**
- If there is a path from (a→b) **or** from (b→a) then they belong to the same **weakly connected component**

CSE 258 – Lecture 12

Web Mining and Recommender Systems

Models of network structure

Network models

A basic problem in network modeling is to define a **random process** that generates networks that are similar to those in the real world

(why?)

- To define a “null model”, i.e., to test assumptions about the properties of the network
- To generate “similar looking” networks with the same properties
- To extrapolate about how a network will look in the future

Network models

The simplest model:

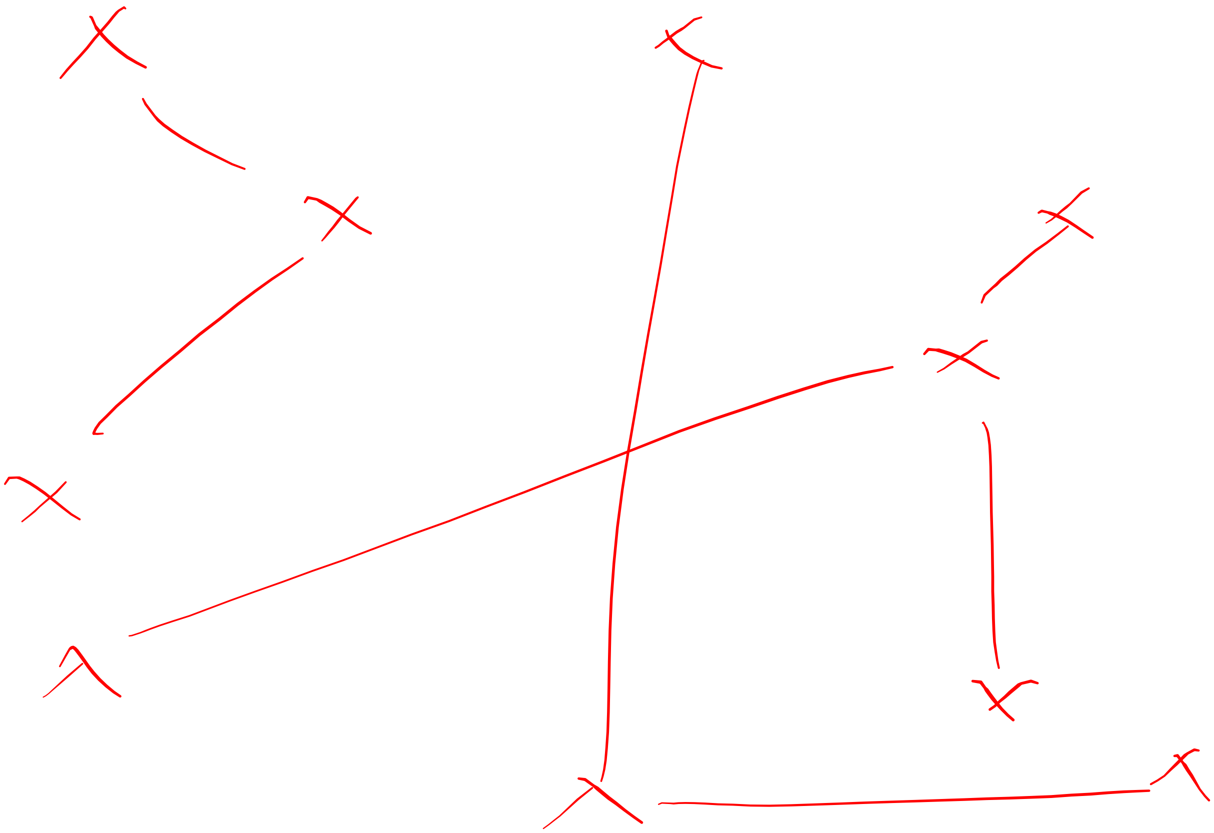
Suppose we want a network with *N* nodes and *E* edges

- Create a graph with *N* nodes
- For every pair of nodes (*i*, *j*), connect them with probability *p*
- If we want the expected number of edges to be *E*, then we should set

$$p = E / \binom{N}{2} \longrightarrow \text{all possible edges}$$

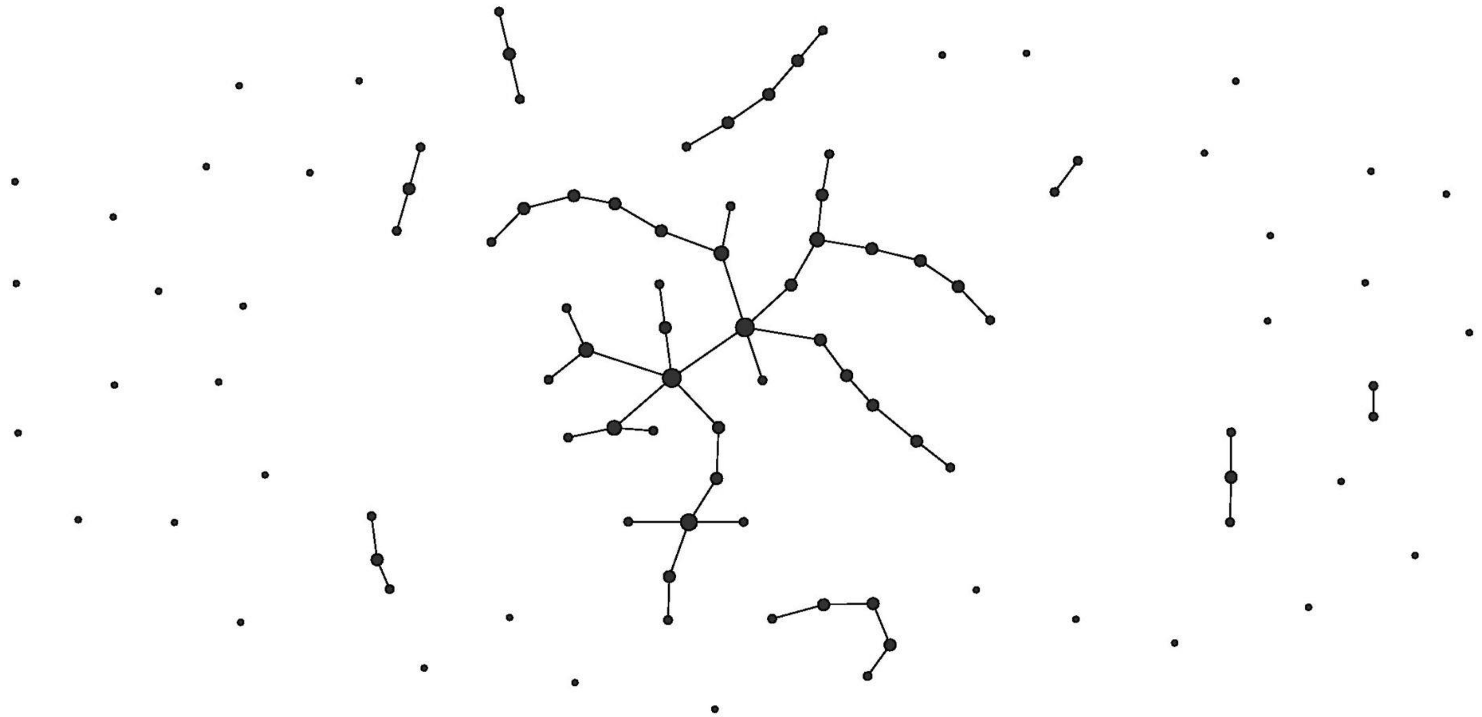
- This is known as the "Erdos-Renyi" random graph model

Network models



Network models

Example of a graph generated by this process (p = 0.01):



(picture from Wikipedia http://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model)

The Erdos-Renyi model

- Do Erdos-Renyi graphs look "realistic"?
- e.g. what sort of degree distributions do they generate, and are those similar to real-world networks?

$$p(\deg(v) = k) =$$

$$\binom{N-1}{k} p^k (1-p)^{N-1-k}$$

The Erdos-Renyi model

$$p(\deg(v) = k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- What does the **degree distribution** of the graph look like as N→infinity, but while (N-1)p remains constant
- In other words, what does the degree distribution converge to if we fix the expected degree = c
  - i.e.:

$$\lim_{N\to\infty} p(\deg(v) = k) = ?$$

Network models

Recall(?): Poisson limit theorem

If $n \to \infty$ and $np \to c$ (with $c > 0$) then

$$\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \to e^{-c} \frac{c^k}{k!}$$

Poisson distribution

proof is "easy": just apply Stirling's approximation for large factorials:

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

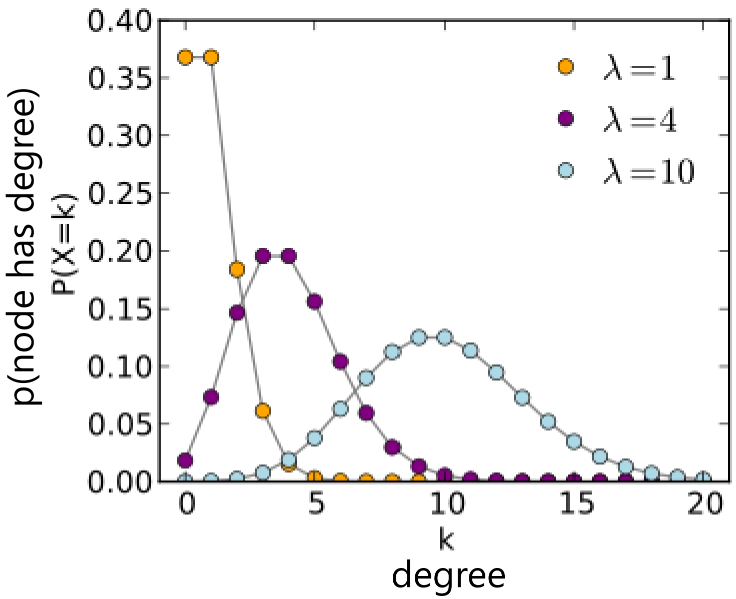
and simplify until you get the desired result

Network models

So, for large graphs, node degrees of an Erdos-Renyi random model are Poisson distributed:

Poisson pmf:

$$\frac{\lambda^k}{k!}e^{-\lambda}$$



Q: But is this actually a realistic degree distribution for real-world networks?

(picture from Wikipedia http://en.wikipedia.org/wiki/Poisson_distribution)

Network models

So, for large graphs, node degrees of an Erdos-Renyi random model are Poisson distributed:

Properties of Erdos-Renyi graphs

(results from Erdos & Renyi's 1960 paper: http://www.renyi.hu/~p_erdos/1960-10.pdf)

expected degree

p→0 for large n

- If np < 1, then the graph will almost surely have no connected components larger than O(log(N))
- If np = 1, then the graph will (almost surely) have a largest connected component of size O(N^{2/3})
- If np is a constant > 1, then the graph will have a single **"giant component"** containing a constant fraction of the vertices. No other component will contain more than O(log(N)) vertices
- Various other obscure properties

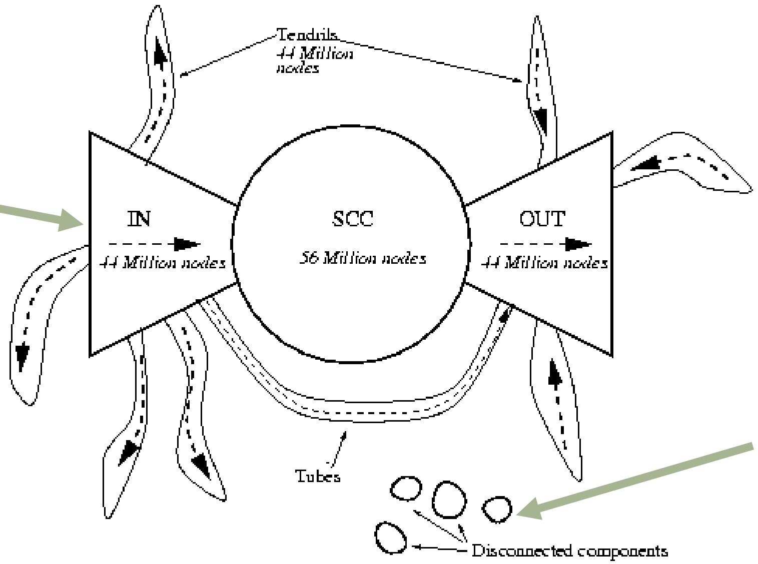
Network models

Which of these results is realistic?

- Giant components

(from Broder et al.'s paper on the structure of the web graph, WWW 2009: http://www9.org/w9cdrom/160/160.html)

Giant strongly connected component



Small disconnected components

(the "bow-tie" and "tentacle" structure of the web)

Which of these results is realistic?

- Giant components

See other examples from the Stanford Network Analysis Collection, e.g.

- astrophysics citation network – 99% of nodes in largest WCC, 37% of nodes in largest SCC
- astrophysics collaboration network – 95% of nodes in largest WCC, 95% of nodes in largest SCC
- Wikipedia talk pages – 99% of nodes in largest WCC, 30% of nodes in largest SCC

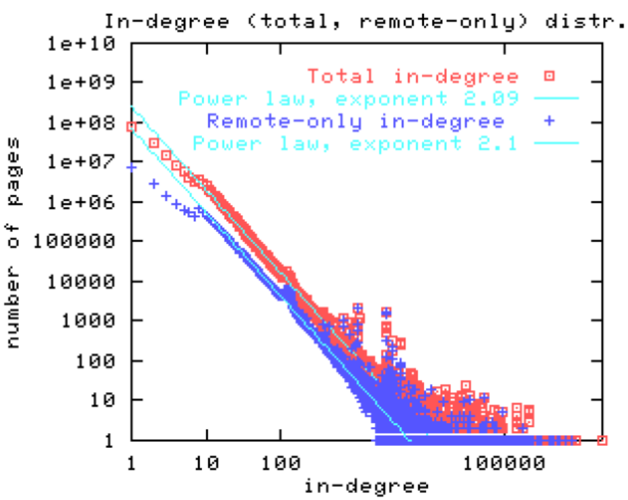
Network models

Which of these results is realistic?

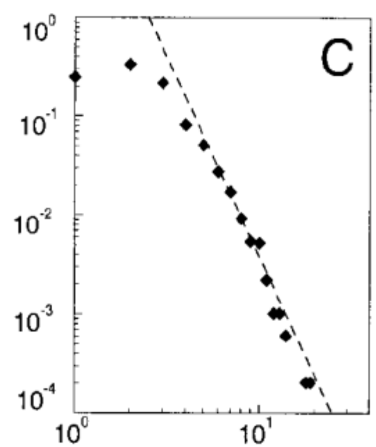
- Poisson-distributed degree distribution?

Degree distributions of a few real-world networks:

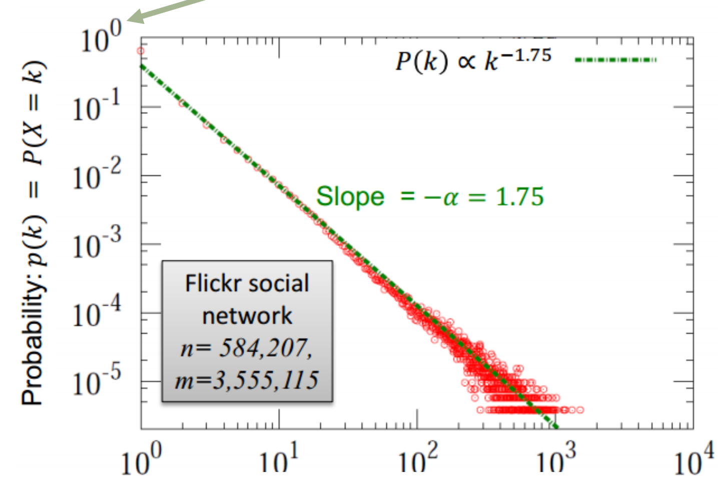
note: log-log plots



e.g. web graph (Broder et al.)



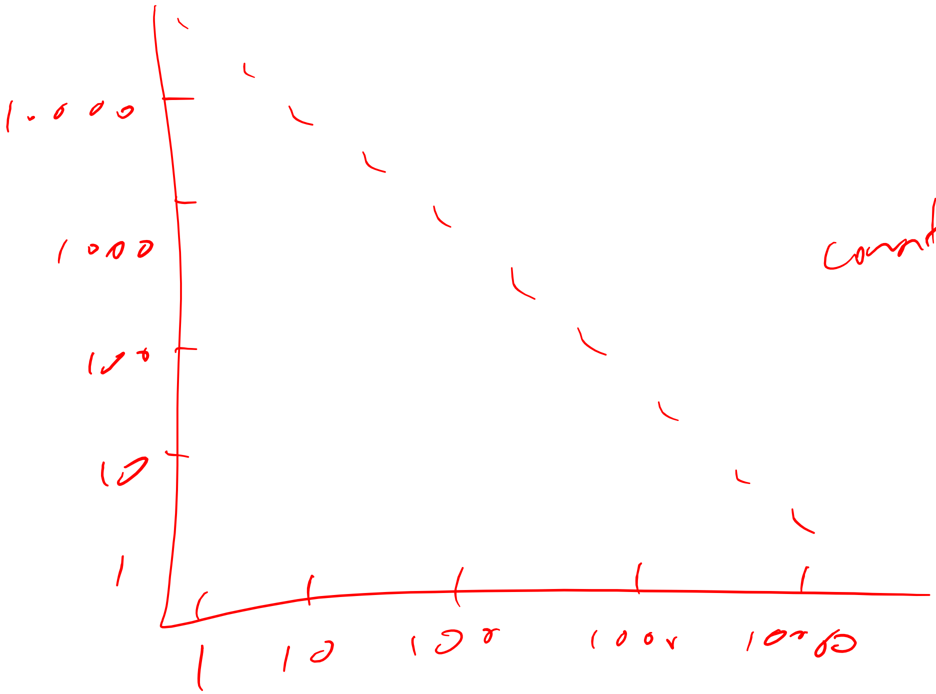
e.g. power grid (Barabasi-Albert)



e.g. Flickr (Leskovec)

Network models

Which of these results is realistic?



Network models

Which of these results is realistic?

- Real-world networks tend to have **power-law** degree distributions

$$p(x) = Cx^{-\alpha}$$

(plotting x against p(x) looks like a straight line on a log-log plot)

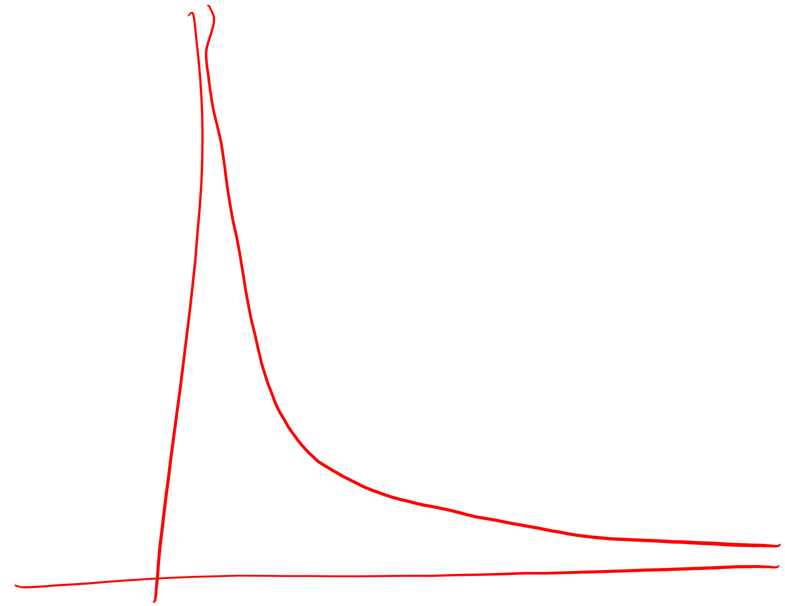
- This is different from a Poisson distribution, which has a mode of *np*

Network models

Which of these results is realistic?

Poisson

Erdos Renyi



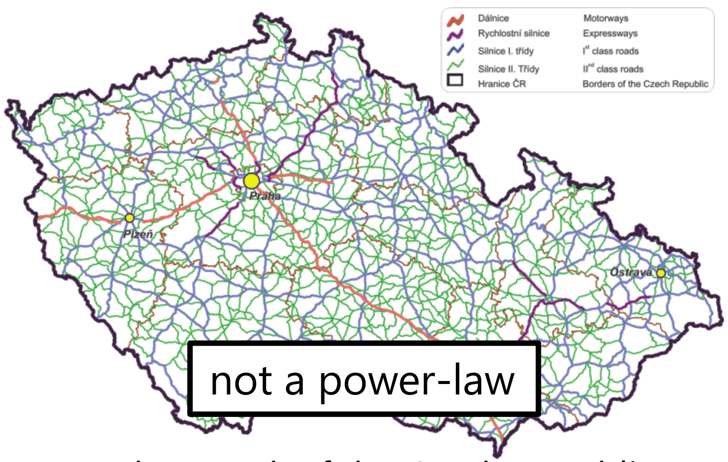
Power law

Real networks

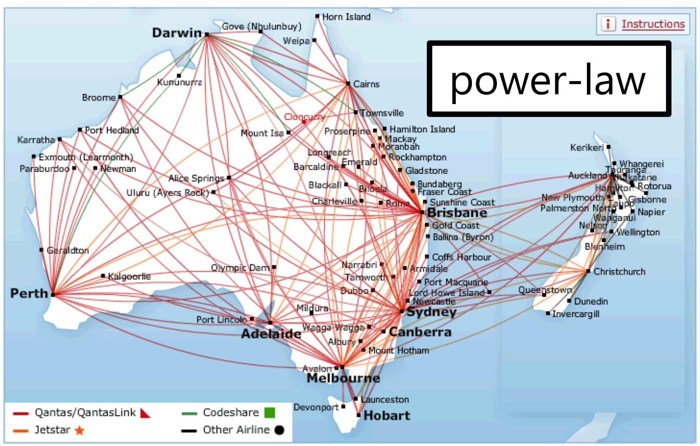
Network models

Which of these results is realistic?

- For example, consider the difference between a **road network** and a **flight network:**



road network of the Czech Republic



Qantas flight network

In the former, nodes have similar degrees; the latter is characterized by a few important “hubs"

(pictures from www.sydos.cz and)

Network models

How can we design a model of network formation that follows a power-law distribution?

- We'd like a model of network formation that produces a small number of "hubs", and a long-tail of nodes with lower degree
- This can be characterized by nodes being more likely to connect to high-degree nodes

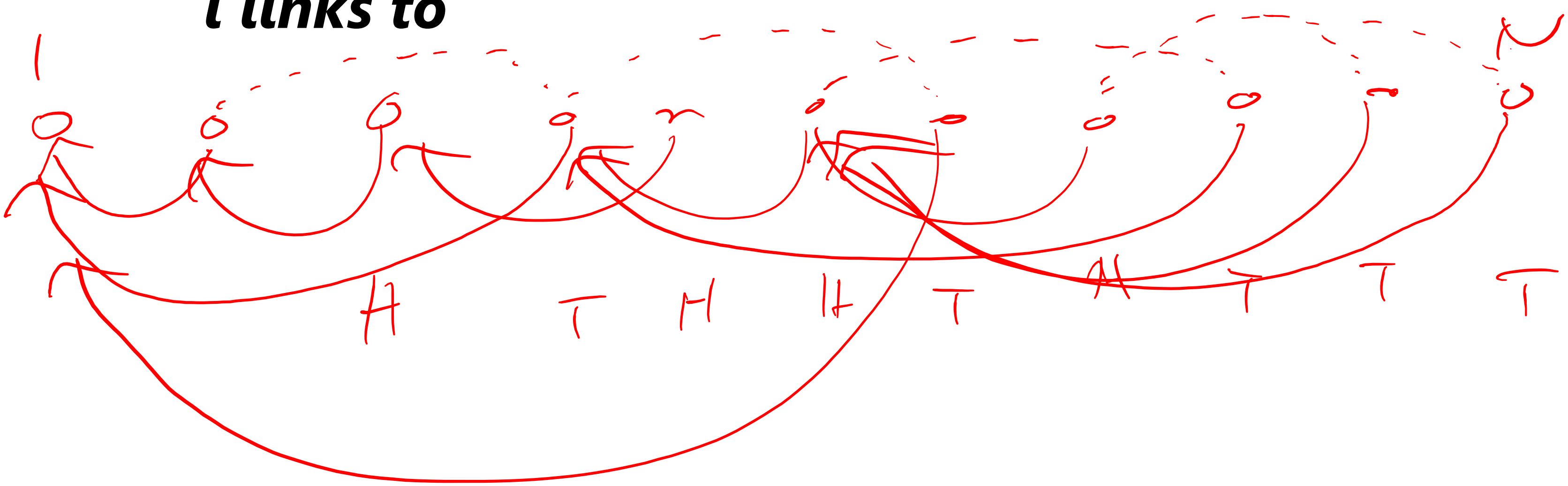
Preferential attachment models of network formation

Consider the following process to generate a network (e.g. a web graph):

1. Order all of the N pages 1,2,3,…,N and repeat the following process for each page *j*:
2. Use the following rule to generate a link to another page:
   a. With probability *p*, link to a random page $i < j$
   b. Otherwise, choose a random page *i* and link to the page ***i links to***

Network models

1. Order all of the N pages 1,2,3,…,N and repeat the following process for each page *j*:
2. Use the following rule to generate a link to another page:
   a. With probability *p*, link to a random page *i* < *j*
   b. Otherwise, choose a random page *i* and link to the page ***i links to***



Preferential attachment models of network formation

- This step is important:

"2b. Choose a random page *i* and link to the page ***i links to***"

- Critically, this will have higher probability of generating links to pages that already have high degree
- It can be rewritten as

"2b. Link to a random page *i* **in proportion to its degree**", i.e.,

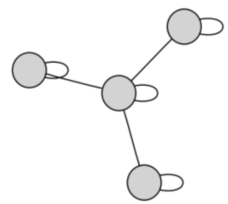
$$p(\text{link to } i) = \frac{\deg(i)}{\sum_j \deg(j)}$$

- This phenomenon is referred to as "rich get richer", i.e., a page that already has many links is likely to get more

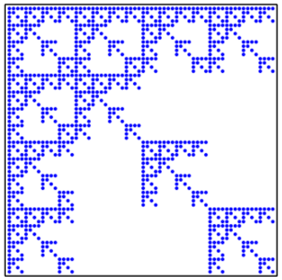
Preferential attachment models of network formation

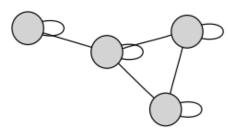
- Most importantly, networks created in this way exhibit power-law distributions (in terms of their **in**-degree) (proof is in Bollobas & Riordan, 2005)
- Specifically, the number of pages with *k* in-links is distributed approximately according to $1/k^c$, where *c* grows as a function of *p* (i.e., the higher the probability that we copy a link from another page, the more likely we are to see extremely popular pages)

Other models of network formation

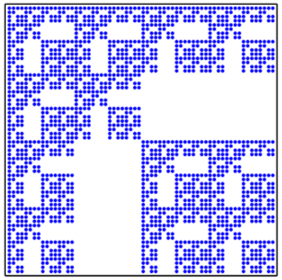


1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	1





1	1	1	1
1	1	0	0
1	0	1	1
1	0	1	1



Initiator K_1 K_1 adjacency matrix K_3 adjacency matrix

- e.g. Kronecker graphs (Leskovec et al., 2010) – are built recursively through Kronecker multiplication of some template
- Intuitively, communities recursively form smaller “copies” of themselves in order to build the complete network

Network models

So far…

- We've seen two models of network formation – Erdos Renyi and Preferential Attachment
- Erdos Renyi captures some of the basic properties of real-world networks (e.g. a single "giant component") but fails to capture power-law distributions, which are ubiquitous in real networks
- Power-law distributions are characterized by the "rich-get-richer" phenomenon – nodes are more likely to connect to other nodes that are already of high degree

"Friendship paradox"

- What are the consequences of a highly imbalanced degree distribution?
- E.g. why does it seem that my friends have more friends than I do?
  - My co-authors have more citations than I do
  - My sexual partners have had more sexual partners than I have
  - etc.

Explanation

Average node degree = $\frac{1}{|V|} \sum_u \deg(u) = \frac{2|E|}{|V|}$

$= \mu$

$\mathrm{var}(\deg(u)) = \frac{1}{|V|} \sum_u \deg(u)^2 - \mu^2$

Explanation

Average degree of a neighbor =

$$\frac{\sum_{\substack{u,v \in E\\ \text{edges}}} \deg(u)}{2|E|}$$

$$= \frac{\sum_u \deg(u)^2}{2|E|}$$

$$= \left(\mathrm{var}(\deg(u)) + \mu^2\right) \times \frac{|U|}{2|E|}$$

$$\frac{1}{\mu}$$

$$= \frac{\mathrm{var}(\deg(u))}{\mu} + \mu$$

$$\mathrm{var}(\deg(u)) = \frac{\sum_u \deg(u)^2}{|U|} - \mu^2$$

Explanation

av. deg $= \mu$

av. deg of friend $= \mu + \dfrac{var(deg)}{\mu}$

Questions?

Further reading:

- Original Erdos-Renyi paper:

"On the evolution of random graphs" (Erdos & Renyi, 1960)

http://www.renyi.hu/~p_erdos/1960-10.pdf

- Power laws:

"Power laws, Pareto distributions and Zipf's law" (Newman, 2005)

http://dx.doi.org/10.1080%2F00107510500052444

- Easley & Kleinberg, Chapter 13 & 18

CSE 258 – Lecture 12

Web Mining and Recommender Systems

Triadic closure; strong & weak ties

Triangles

So far we've seen (a little about) how networks can be characterized by their connectivity patterns

What more can we learn by looking at higher-order properties, such as relationships between **triplets** of nodes?

Motivation

Q: Last time you found a job, was it through:

- A complete stranger?
- A close friend?
- An acquaintance?

A: Surprisingly, people often find jobs through **acquaintances** rather than through close friends (Granovetter, 1973)

Motivation

- Your friends (hopefully) would seem to have the greatest motivation to help you
- But! Your closest friends have limited information that you don't already know about
- Alternately, acquaintances act as a "bridge" to a different part of the social network, and expose you to new information

This phenomenon is known as **the strength of weak ties**

Motivation

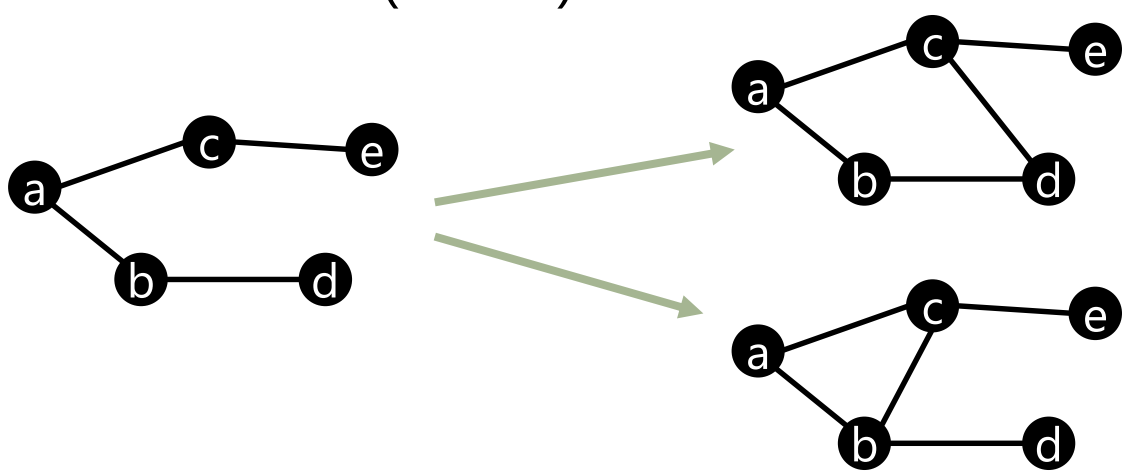
- To make this concrete, we'd like to come up with some notion of "tie strength" in networks
- To do this, we need to go beyond just looking at edges in isolation, and looking at how an edge connects one part of a network to another

Refs:
"The Strength of Weak Ties", Granovetter (1973): http://goo.gl/wVJVlN
"Getting a Job", Granovetter (1974)

Triangles

Triadic closure

Q: Which edge is most likely to form **next** in this (social) network?



(a)

(b)

A: (b), because it creates a **triad** in the network

Triangles

"If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future" (Ropoport, 1953)

Three reasons (from Heider, 1958; see Easley & Kleinberg):

- Every mutual friend *a* between *bob* and *chris* gives them an **opportunity** to meet
- If *bob* is friends with *ashton*, then knowing that *chris* is friends with *ashton* gives *bob* a reason to **trust** *chris*
- If *chris* and *bob* don't become friends, this causes stress for *ashton* (having two friends who don't like each other), so there is an **incentive** for them to connect

Triangles

The extent to which this is true is measured by the (local) **clustering coefficient:**

- The clustering coefficient of a node *i* is the probability that two of *i*'s friends will be friends with each other:

neighbours of i

pairs of neighbours that are edges

$$C_i = \frac{\sum_{j,k \in \Gamma(i)} \delta((j,k) \in E)}{k_i(k_i - 1)}$$

(edges (j,k) and (k,j) are both counted for undirected graphs)

degree of node i

- This ranges between 0 (none of my friends are friends with each other) and 1 (all of my friends are friends with each other)

Triangles

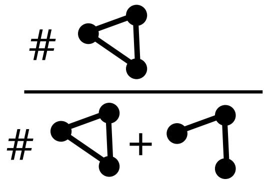
The extent to which this is true is measured by the (local) **clustering coefficient:**

- The clustering coefficient of the **graph** is usually defined as the average of local clustering coefficients

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i$$

- Alternately it can be defined as the fraction of connected triplets in the graph that are closed (these do not evaluate to the same thing!):

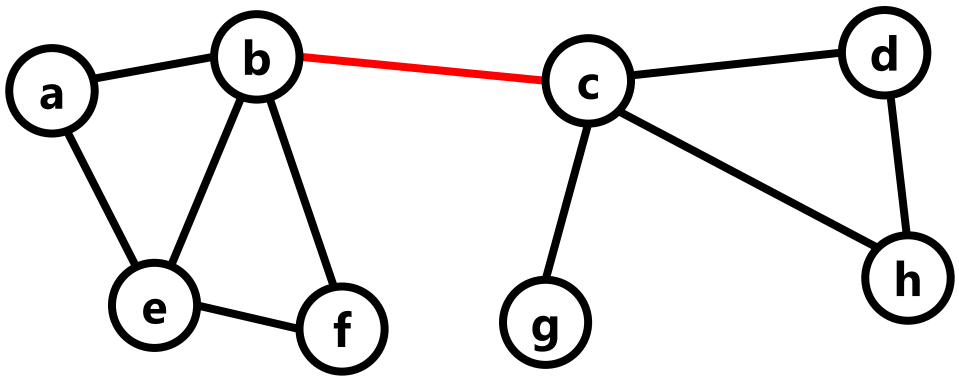
$$C = \frac{\# \text{ of closed triplets}}{\# \text{ of connected triplets}}$$



Bridges

Next, we can talk about the role of edges in relation to the rest of the network, starting with a few more definitions

1. Bridge edge

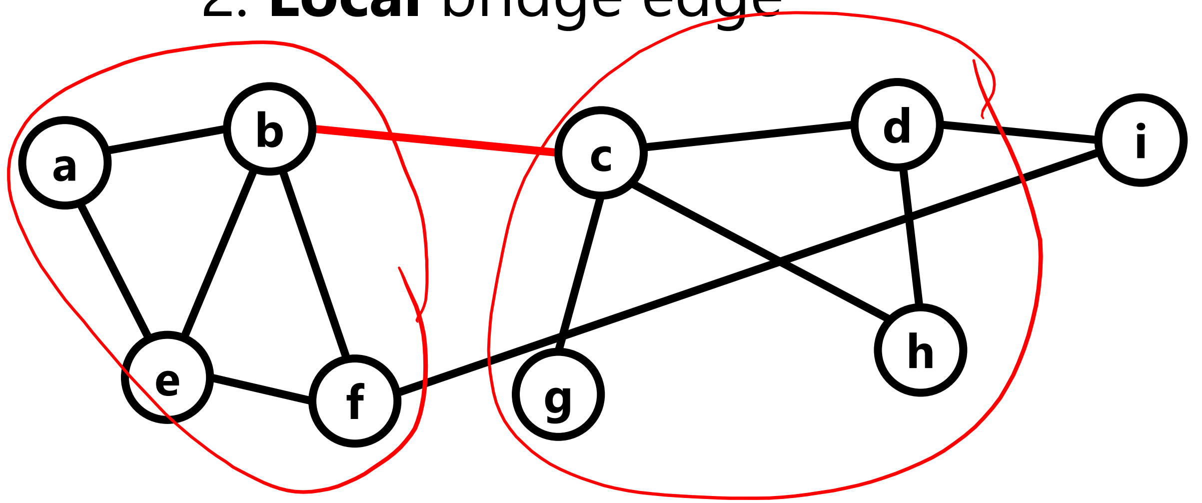


An edge (b,c) is a **bridge** edge if removing it would leave no path between b and c in the resulting network

Bridges

In practice, “bridges” aren’t a very useful definition, since there will be very few edges that completely isolate two parts of the graph

2. **Local** bridge edge

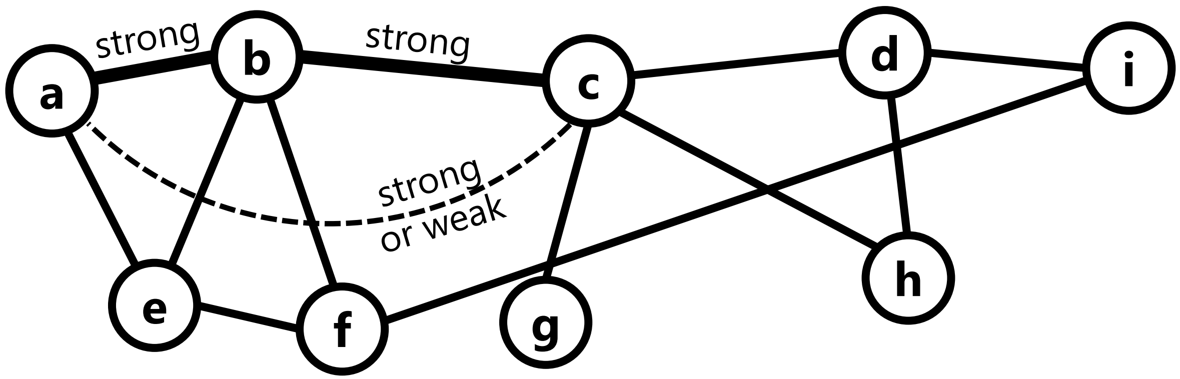


An edge (b,c) is a **local bridge** if removing it would leave no edge between b’s friends and c’s friends (though there could be more distant connections)

Strong & weak ties

We can now define the concept of “strong” and “weak” ties (which roughly correspond to notions of “friends” and “acquaintances”

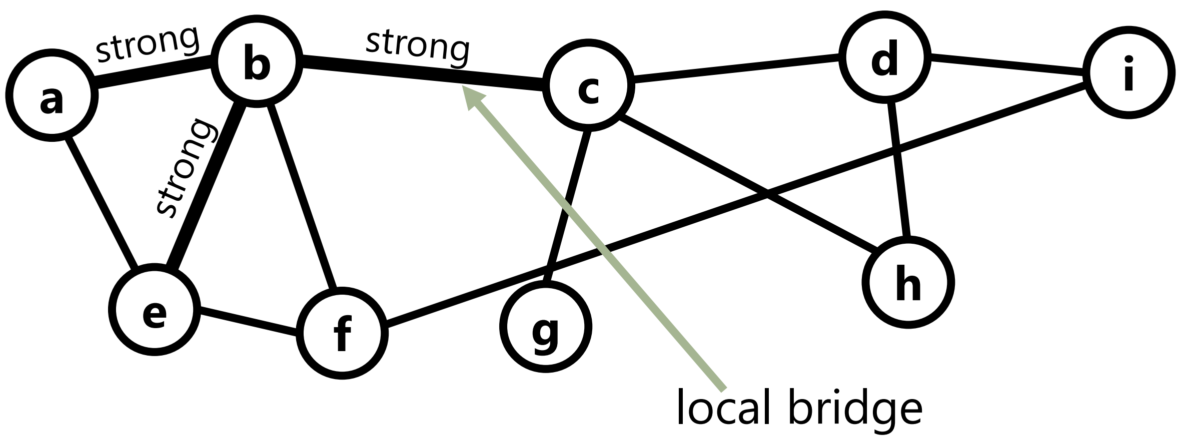
3. Strong triadic closure property



If (a,b) and (b,c) are connected by **strong** ties, there must be at least a **weak** tie between a and c

Strong & weak ties

Granovetter's theorem: if the strong triadic closure property is satisfied for a node, and that node is involved in two strong ties, then any incident local bridge must be a **weak tie**



Proof (by contradiction): (1) b has two strong ties (to a and e); (2) suppose it has a **strong** tie to c via a local bridge; (3) but now a tie must exist between c and a (or c and e) due to strong triadic closure; (4) so b → c cannot be a bridge

Strong & weak ties

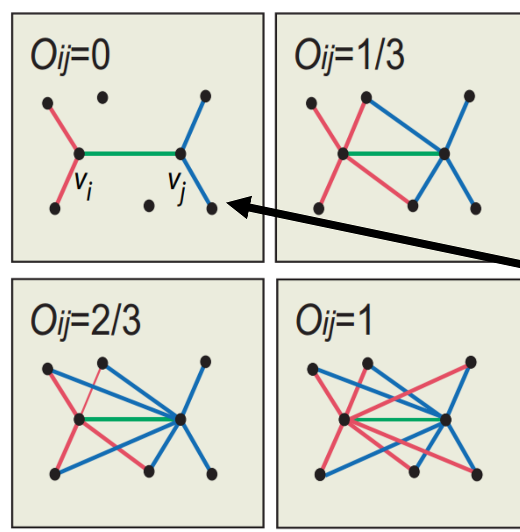
Granovetter's theorem: so, if we're receiving information from distant parts of the network (i.e., via "local bridges") then we must be receiving it via **weak ties**

Q: How to test this theorem empirically on real data?
A: Onnela et al. 2007 studied networks of mobile phone calls

Defn. 1: Define the "overlap" between two nodes to be the Jaccard similarity between their connections

$$O_{i,j} = \frac{\Gamma(i) \cap \Gamma(j)}{\Gamma(i) \cup \Gamma(j)}$$

neighbours of i



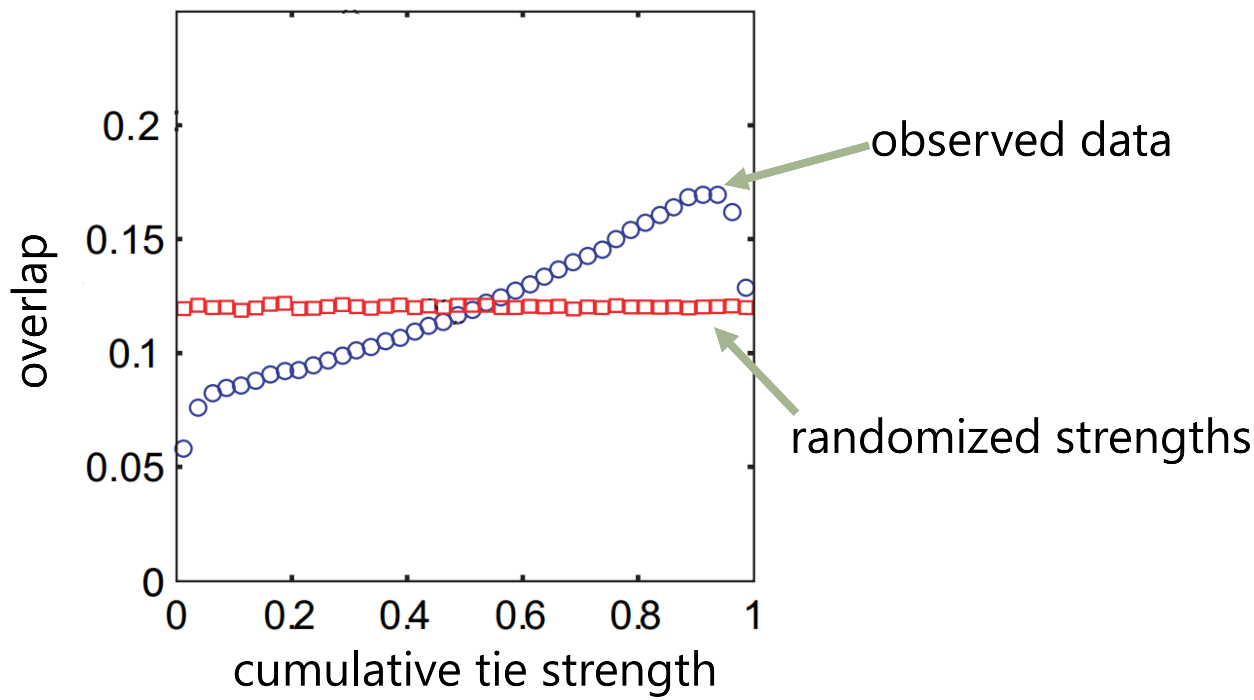
"local bridges" have overlap 0

(picture from Onnela et al., 2007)

Strong & weak ties

Secondly, define the "strength" of a tie in terms of the number of phone calls between *i* and *j*

finding: the "stronger" our tie, the more likely there are to be additional ties between our mutual friends



(picture from Onnela et al., 2007)

Strong & weak ties

Another case study (Ugander et al., 2012)

Suppose a user receives four e-mail invites to join facebook from users who are already on facebook. Under what conditions are we most likely to accept the invite (and join facebook)?

1. If those four invites are from four close friends?
2. If our invites are from found acquiantances?
3. If the invites are from a combination of friends, acquaintances, work colleagues, and family members?

hypothesis: the invitations are most likely to be adopted if they come from **distinct groups** of people in the network

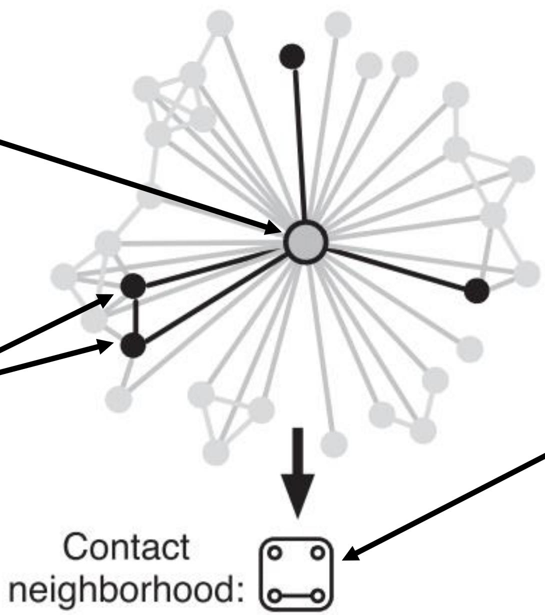
Strong & weak ties

Another case study (Ugander et al., 2012)

Let's consider the connectivity patterns amongst the people who tried to recruit us

user being recruited

users recruiting



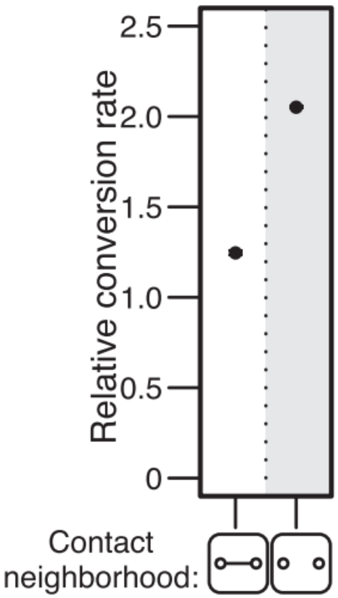
reachability between users attempting to recruit

(picture from Ugander et al., 2012)

Strong & weak ties

Another case study (Ugander et al., 2012)

Let's consider the connectivity patterns amongst the people who tried to recruit us



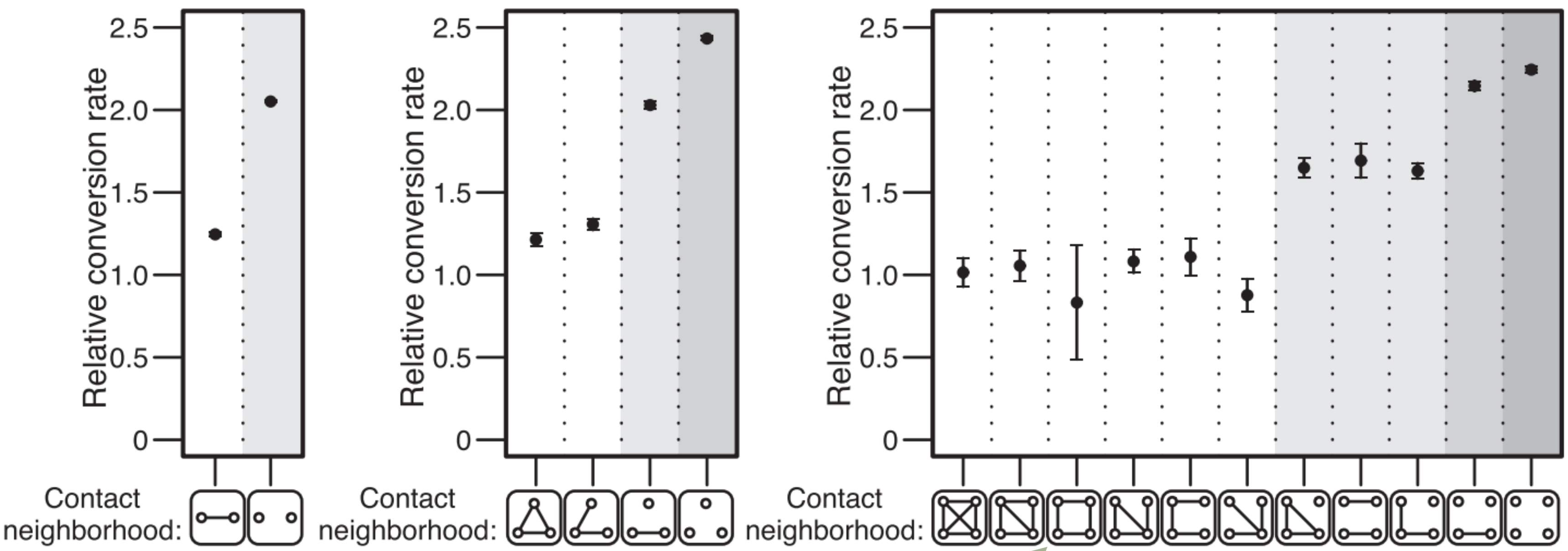
- **Case 1:** two users attempted to recruit
- **y-axis:** relative to recruitment by a single user
- **finding:** recruitments are **more likely to succeed** if they come from friends who are **not connected to each other**

(picture from Ugander et al., 2012)

Strong & weak ties

Another case study (Ugander et al., 2012)

Let's consider the connectivity patterns amongst the people who tried to recruit us



error bars are high since this structure is very very rare

(picture from Ugander et al., 2012)

Strong & weak ties

So far:

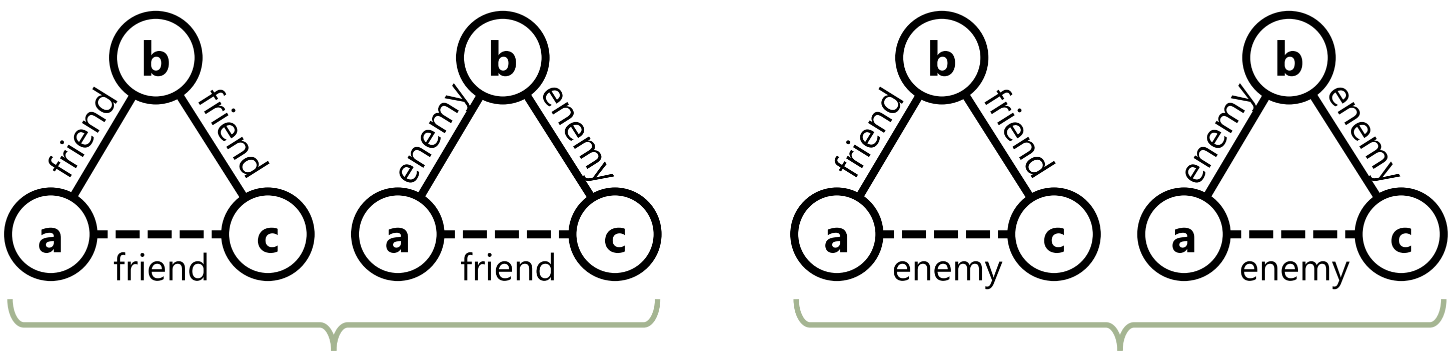
Important aspects of network structure can be explained by the way an edge connects two parts of the network to each other:

- Edges tend to close open triads (clustering coefficient etc.)
- It can be argued that edges that bridge different parts of the network somehow correspond to “weak” connections (Granovetter; Onnela et al.)
- Disconnected parts of the networks (or parts connected by local bridges) expose us to distinct sources of information (Granovettor; Ugander et al.)

See also…

Structural balance

Some of the assumptions that we've seen today may not hold if edges have **signs** associated with them



balanced: the edge a→c is **likely** to form

imbalanced: the edge a→c is **unlikely** to form

(see e.g. Heider, 1946)

Questions?

Further reading:

- Easley & Kleinberg, Chapter 3
- The strength of weak ties (Granovetter, 1973) http://goo.gl/wVJVlN
- Bearman & Moody "Suicide and friendships among American adolescents" http://www.soc.duke.edu/~jmoody77/suicide_ajph.pdf
- Onnela et al.'s mobile phone study "Structure and tie strengths in mobile communication networks" http://www.hks.harvard.edu/davidlazer/files/papers/Lazer_PNAS_2007.pdf
- Ugander et al.'s facebook study "Structural diversity in social contagion" file:///C:/Users/julian/Downloads/PNAS-2012-Ugander-5962-6.pdf