

# CMSC5719 Introductions to Social Computing

Irwin King

The Chinese University of Hong Kong

<http://www.cse.cuhk.edu.hk/irwin.king/confs/cmsc5719>  
[king@cse.cuhk.edu.hk](mailto:king@cse.cuhk.edu.hk)

©2012 All Rights Reserved.

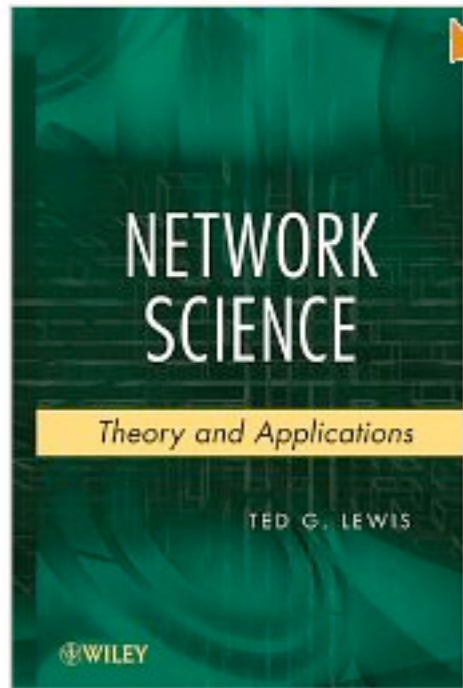


# Course Information

- Irwin King
- 908 Ho Sin Hang Engineering Building
- <http://www.cse.cuhk.edu.hk/irwin.king>
- email: [king@cse.cuhk.edu.hk](mailto:king@cse.cuhk.edu.hk)
- <http://www.cse.cuhk.edu.hk/irwin.king/confs/cm5719>
- Office hours: M8 and T8 from 3:30 pm - 4:30 pm



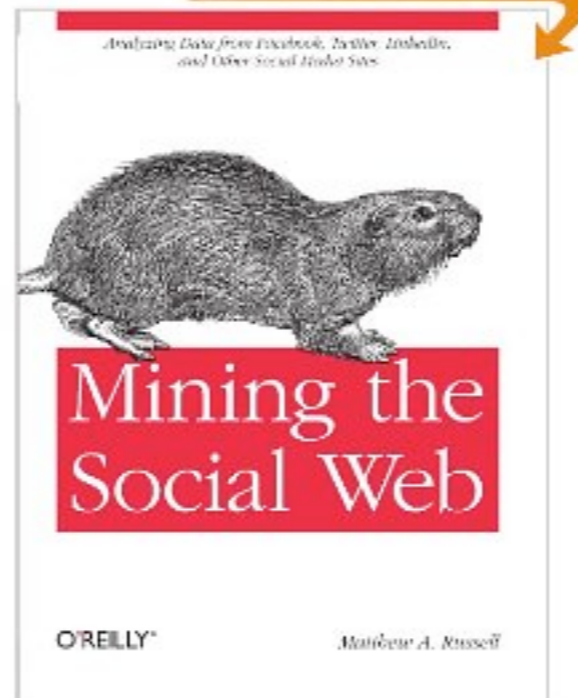
Click to **LOOK INSIDE!**



Network Science:  
Theory and Applications

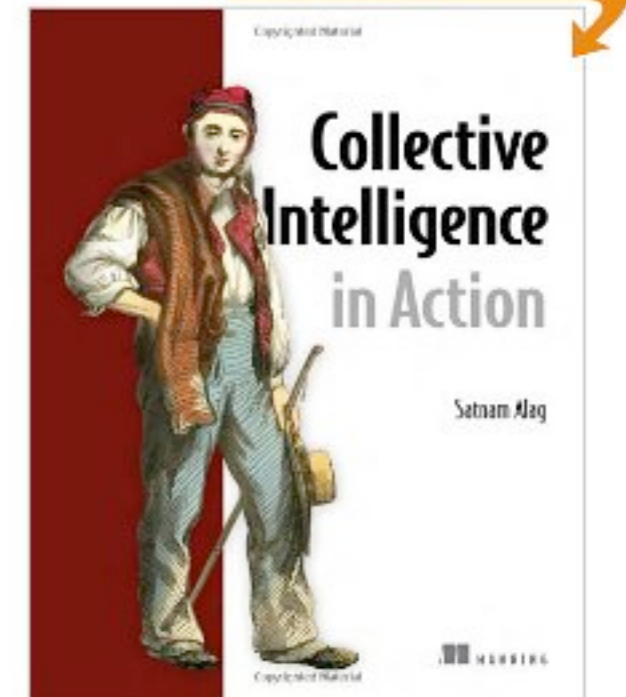
# References

Click to **LOOK INSIDE!**



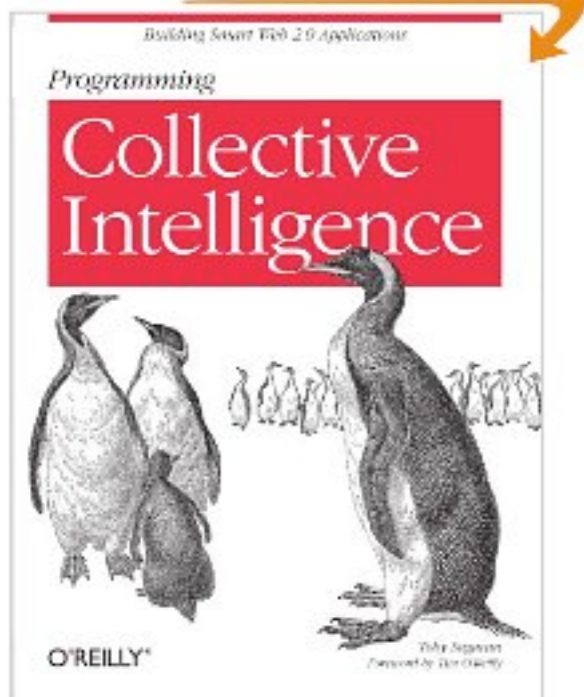
Mining the Social Web:  
Analyzing Data from  
Facebook, Twitter,  
LinkedIn, and Other Social  
Media Sites

Click to **LOOK INSIDE!**



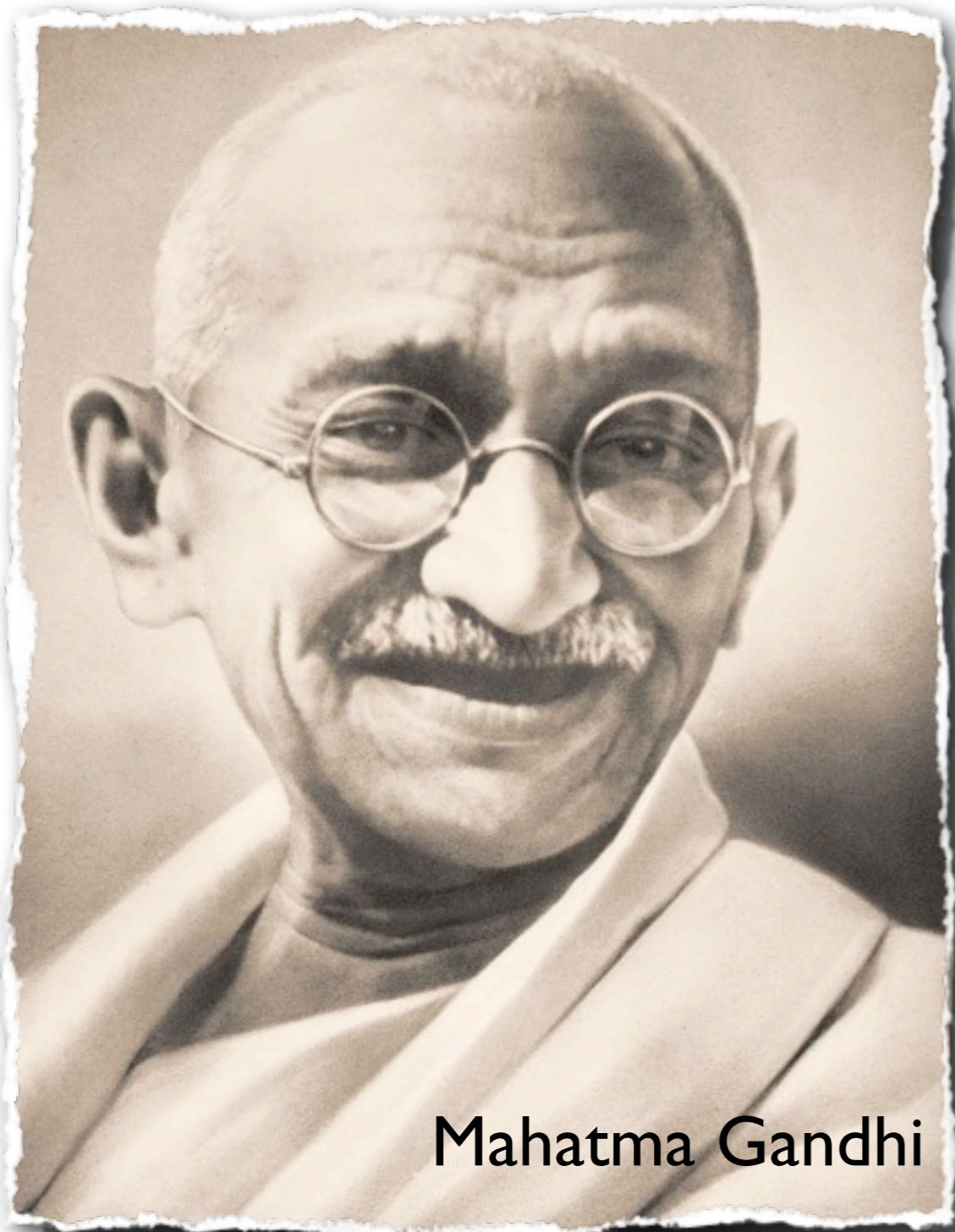
Collective Intelligence in  
Action

Click to **LOOK INSIDE!**



Programming Collective Intelligence:  
Building Smart Web 2.0 Applications





Mahatma Gandhi

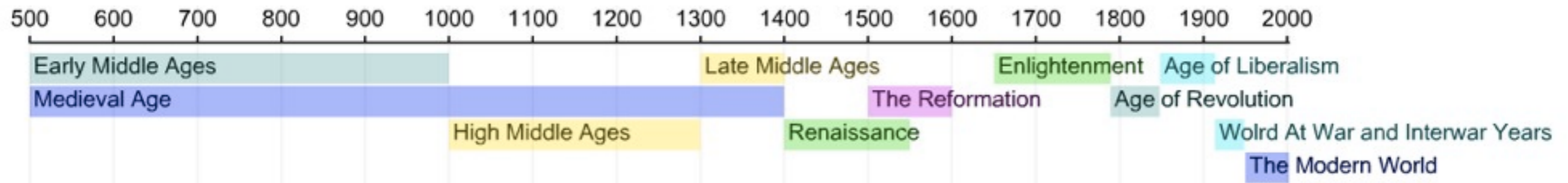
*Interdependence is and ought to be as much the ideal of man as self-sufficiency.*

*Man is a social being.*

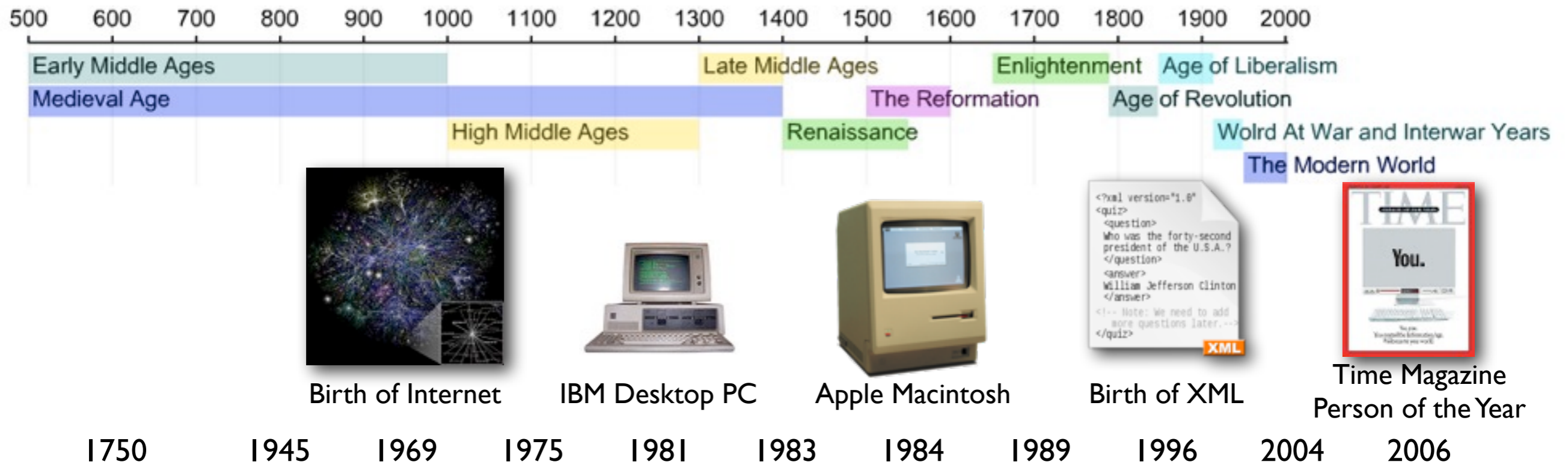




# A Brief History of the World



# A Brief History of the World



**Industrial Revolution**

**Information Age**

**Internet Age**

**www Age**

**Attention Age**

ENIAC



The MITS Altair  
Apple II



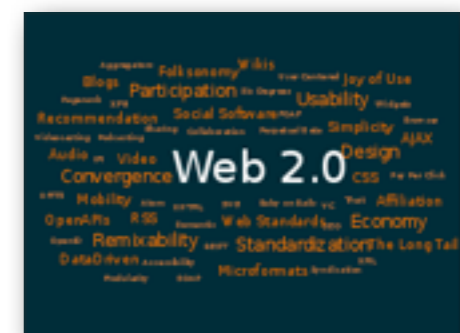
Time Magazine  
Person of the Year



Birth of WWW



Birth of Web 2.0







# revolution in evolution

Highlights from the Journey to 1 Billion PCs

1,000,000,000  
900,000,000  
800,000,000  
700,000,000  
600,000,000  
500,000,000  
400,000,000  
300,000,000  
200,000,000  
100,000,000

1,000,000,000  
900,000,000  
800,000,000  
700,000,000  
600,000,000  
500,000,000  
400,000,000  
300,000,000  
200,000,000  
100,000,000



**1971** - Intel, founded by Robert Noyce, Gordon Moore and Andy Grove, introduces the world's first microprocessor and calls it the Intel® 4004.

**1974** - Intel introduces the 8008 microprocessor, which was used in the first commercially successful personal computer - the Intel 8080.

**1976** - Apple Computer, Inc. releases the Apple I, the first single-circuit board computer. The following year, the company introduces the Apple II, a first for a personal computer, the Apple II featured color graphics.



Microsoft ships the Windows® operating system with a graphical user interface.

America Online is founded.

**1986** - The number of PCs shipped worldwide reaches nearly 64 million and a 15-year period of continuous growth begins.

**1987** - Toshiba introduces the T1000 laptop PC, making portable computing more widely available.

**1988** - Recordable CD discs become available.



**1991** - Creative Labs introduces a Multimedia Upgrade Kit containing a CD-ROM drive, Sound Blaster® Pro card, speakers and multimedia software.

**1992** - The number of PCs shipped worldwide reaches 200 million.

**1993** - The number of PCs shipped worldwide reaches 347 million.



www.intel.com  
© Intel Corporation  
All rights reserved.

\*Source: Intel  
\*\*Source: Intel  
\*\*\*Source: Intel  
\*\*\*\*Source: Intel  
\*\*\*\*\*Source: Intel

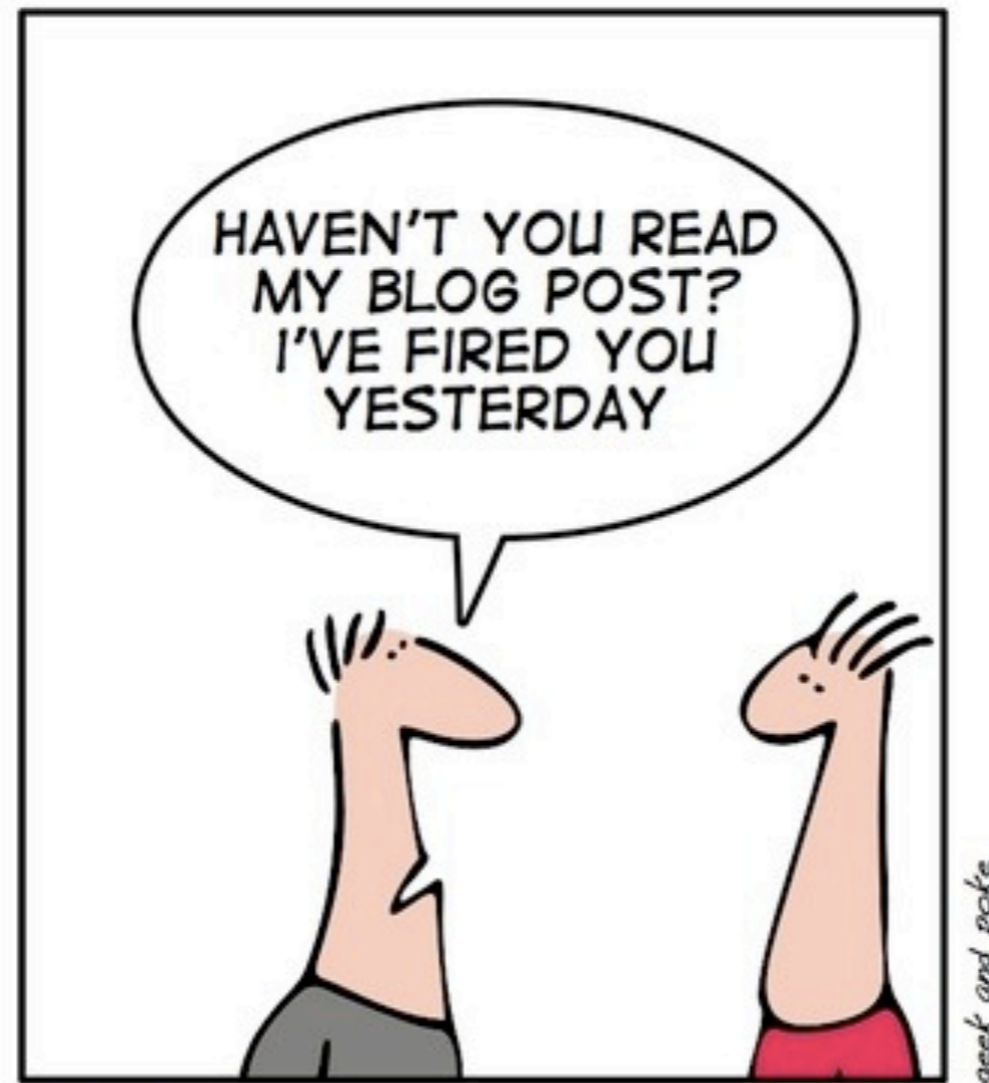
For more information, please visit <http://www.intel.com>





# Social Networking

*HOW TO USE WEB 2.0 IN THE ENTERPRISE*



*PART 1:  
COMMUNICATE WITH YOUR EMPLOYEES*





# Billionaires' Shuffle

2007



Facebook in 2004.02

**2008**

at **23** and \$ **1.5** billion later...

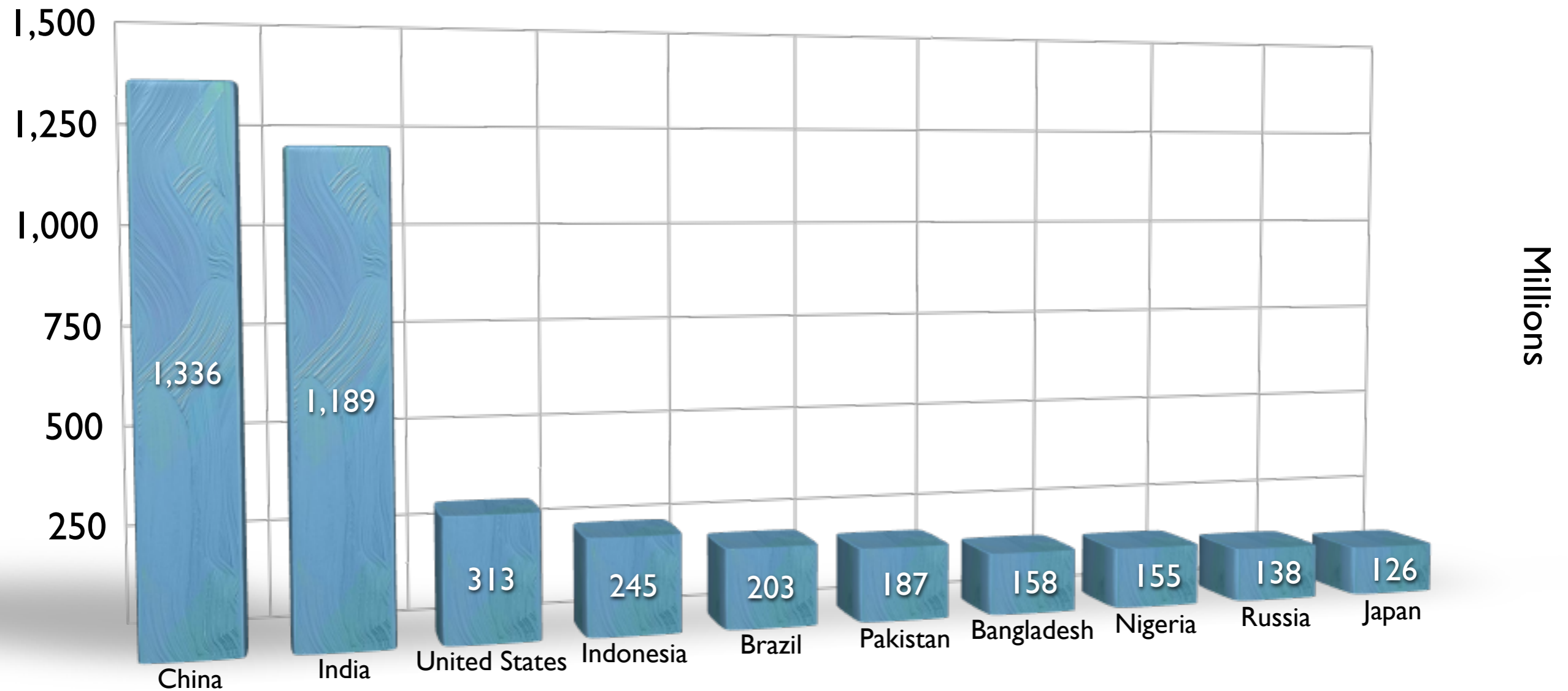


2008



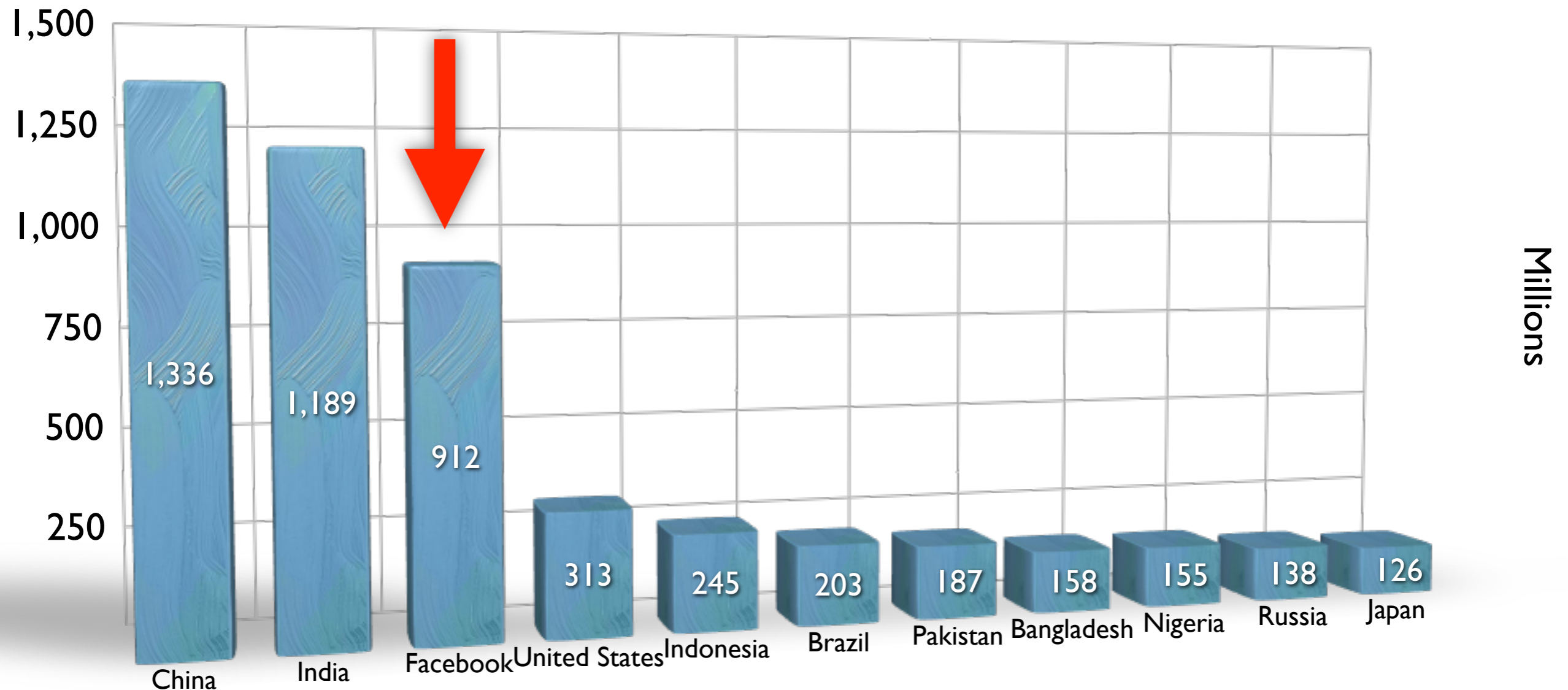
# Top 10 Populations by Countries

as of December 31, 2011



# Top 10 Populations by Countries

as of September 8, 2012





# Facebook's Global Audience

Global Audience: 316,402,840

Data for 11/03/2009

About CheckFacebook.com

[Ads by Google](#)

[Facebook](#)

[Social Search](#)

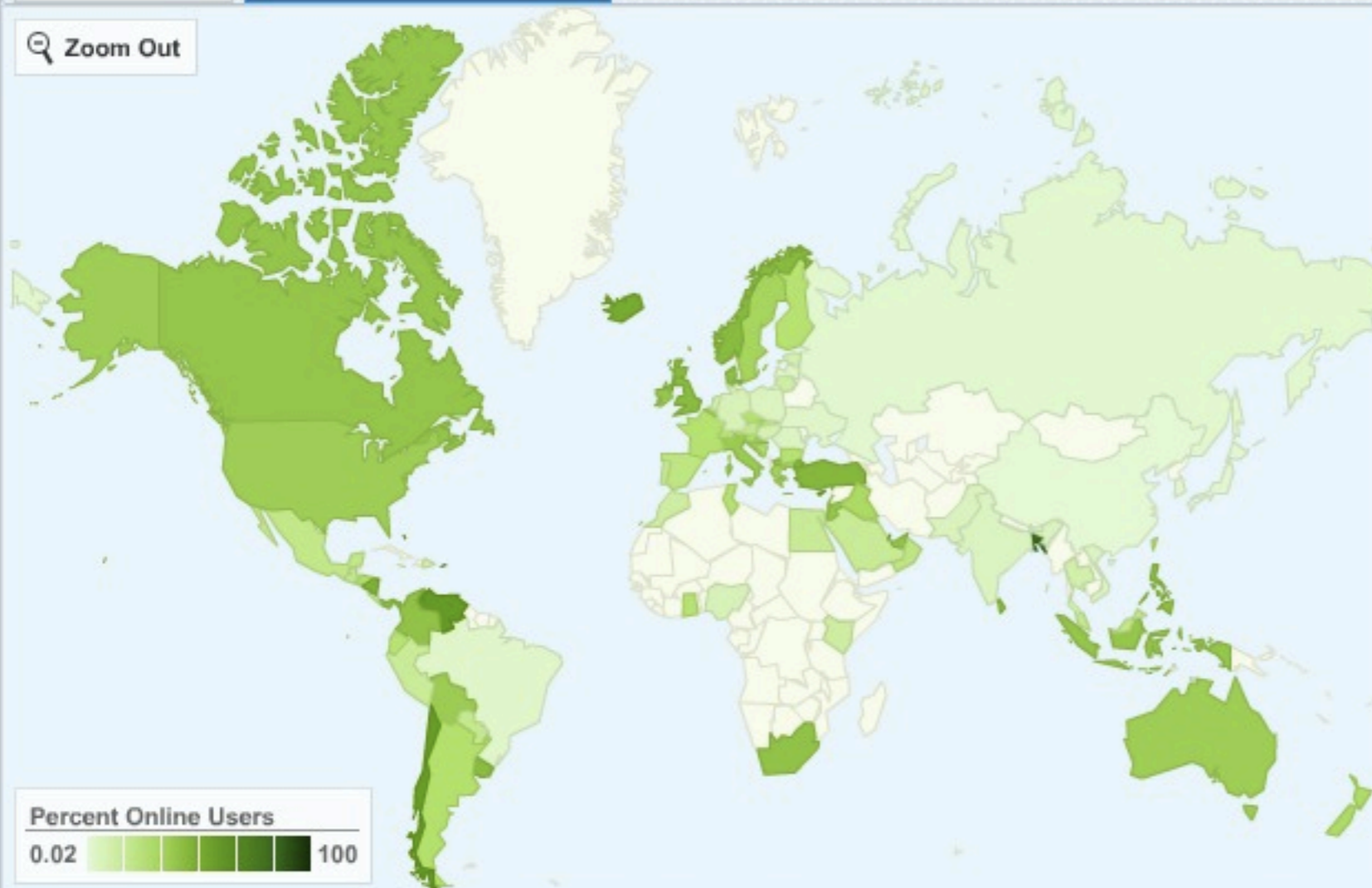
[Twitter](#)

[Blog Marketing](#)

Total Users

% Online Population

Zoom Out



Percent Online Users

0.02 100

Not Pictured: [Hong Kong](#), [Maldives](#), [Palestine](#), [Singapore](#), [Taiwan](#)

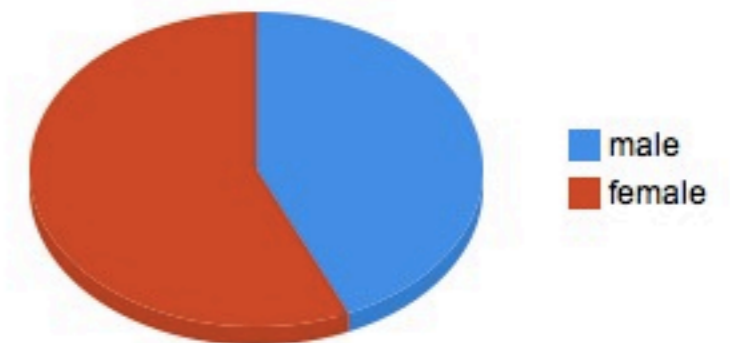
## United States

Country Audience: 94,748,820

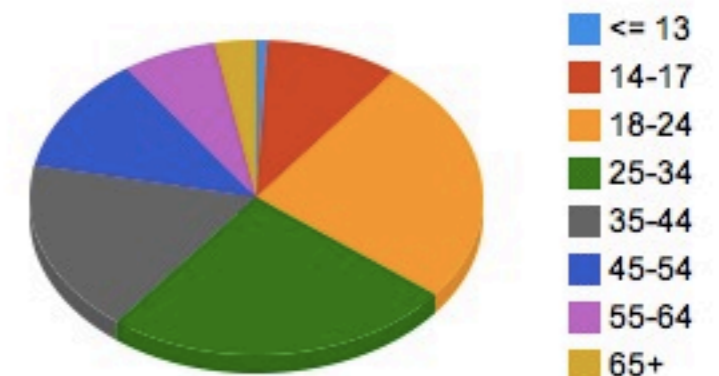
Percent of Global Audience: 29.95%

Share This Site 1543 retweet

United States Male / Female



United States Age Distribution





# Facebook's Global Audience

Global Audience: 912,496,580

Data for 09/08/2012

About CheckFacebook.com



Start monitoring your Facebook Page today!

- ✓ Analyze your competition
- ✓ Track most engaging content
- ✓ Compare your results with leaders

Most comprehensive analytics platform for Facebook out there

Start right NOW!

George May | DOB

## China

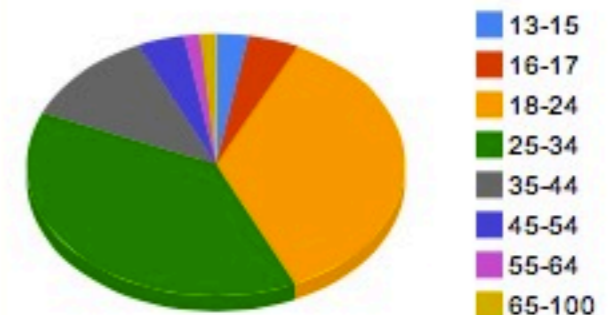
Country Audience: 552,560  
Percent of Global Audience: 0.13%

Share This Site 4163 retweet

China Male / Female

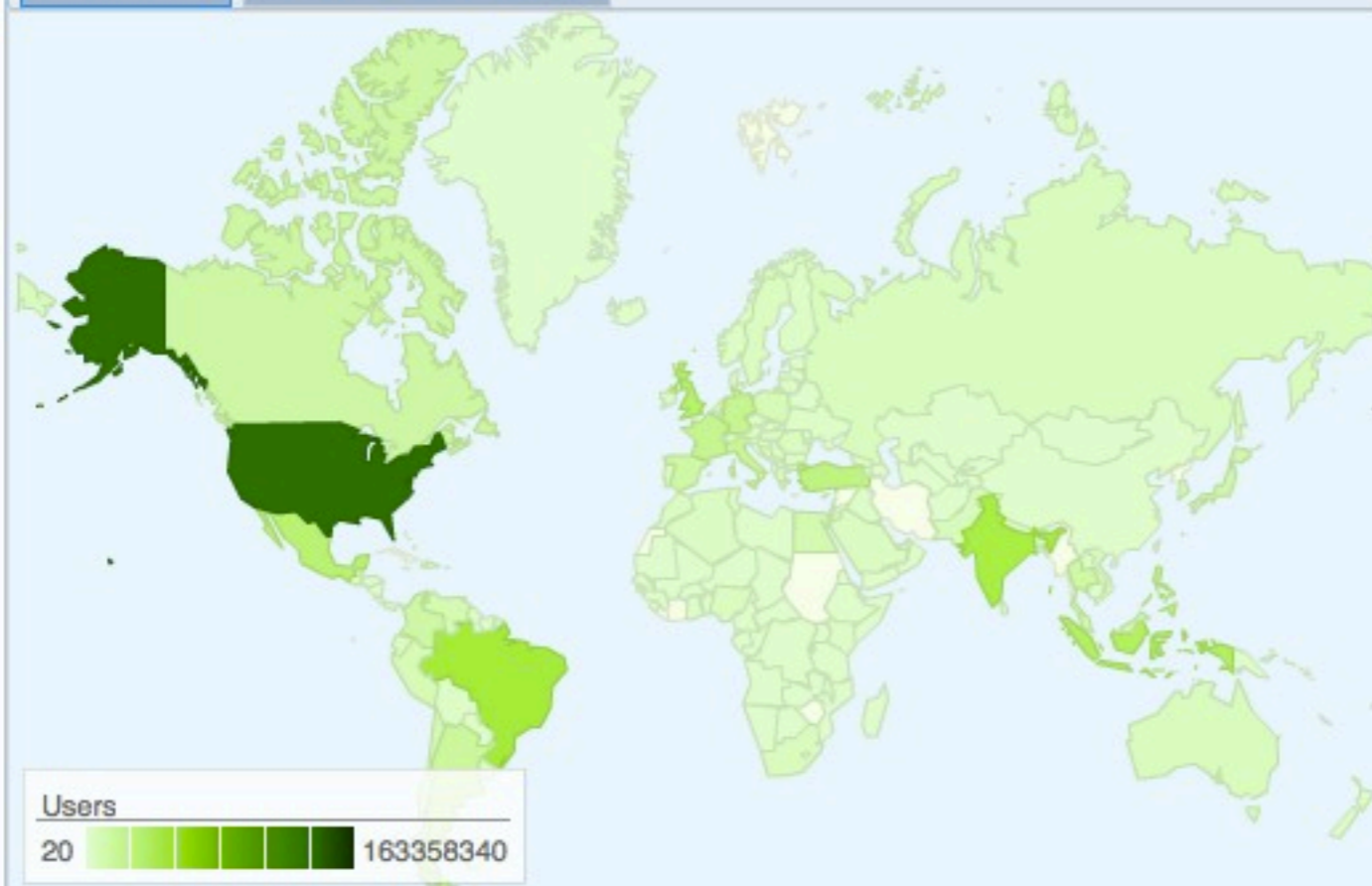


China Age distribution



Total Users

% Online Population



Users

20 163358340

Not Pictured: Hong Kong, Maldives, Palestine, Singapore, Taiwan

PRO Analytics  
for facebook page!

- ✓ Local Industry benchmarks
- ✓ Discover what works
- ✓ Competitive reporting



# Facebook's Growth Stats

## Statistics

### Company Figures

More than 400 million active users  
50% of our active users log on to Facebook in any given day  
More than 35 million users update their status each day  
More than 60 million status updates posted each day  
More than 3 billion photos uploaded to the site each month  
More than 5 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) shared each week

### 10 Largest Countries

1. United States	94,748,820
2. United Kingdom	22,261,080
3. Turkey	14,215,880
4. France	13,396,760
5. Canada	13,228,380
6. Italy	12,581,060
7. Indonesia	11,759,980
8. Spain	7,313,160
9. Australia	7,176,640
10. Philippines	6,991,040

### 10 Fastest Growing Over Past Week

1. Poland	12.46 %	137,900
2. Thailand	10.96 %	161,300
3. Portugal	9.81 %	80,040
4. South Africa	9.25 %	189,080
5. Taiwan	7.82 %	367,400
6. Romania	7.65 %	28,060
7. Germany	7.54 %	350,240
8. Malaysia	7.43 %	236,840
9. Indonesia	6.84 %	752,640
10. Iraq	6.72 %	6,380





# Facebook's Growth Stats

(as of September 2012)

## Statistics

955 million monthly active users at the end of June 2012.

Approximately 81% of our monthly active users are outside the U.S. and Canada.

552 million daily active users on average in June 2012.

543 million monthly active users who used Facebook mobile products in June 2012.

10 Largest Countries		10 Fastest Growing Over Past Week	
1. United States	163,358,340	1. Vietnam	100.09 % 3,598,480
2. Brazil	56,804,900	2. Brazil	0.18 % 100,060
3. India	53,624,320	3. Thailand	0.21 % 34,780
4. Indonesia	44,156,440	4. Colombia	0.19 % 32,060
5. United Kingdom	40,036,380	5. Romania	0.44 % 21,940
6. Mexico	37,542,740	6. Croatia	0.84 % 13,120
7. Turkey	31,108,760	7. Netherlands	12.76 % 9,720
8. Philippines	29,136,040	Antilles	
9. France	24,639,540	8. Canada	0.05 % 9,700
10. Germany	24,300,340	9. Chile	0.10 % 9,560
		10. Jordan	0.37 % 8,980



# Global Internet Traffic

Alexa as of August 2011	China	USA	Japan	India	Brazil	Global
1	Baidu	Google	Yahoo.jp	Google.in	Google.br	Google
2	<b>QQ</b>	<b>Facebook</b>	Google.jp	Google	Google	<b>Facebook</b>
3	Sina	Yahoo!	<b>FC2</b>	<b>Facebook</b>	<b>Facebook</b>	<b>YouTUBE</b>
4	Taobao	<b>YouTUBE</b>	<b>YouTUBE</b>	<b>YouTUBE</b>	<b>YouTUBE</b>	Yahoo!
5	Google.hk	Amazon	Google	Yahoo!	Universo Online	<b>Blogger</b>
6	163	<b>Wikipedia</b>	<b>Ameblo.jp</b>	<b>Blogger</b>	Windows Live	Baidu
7	<b>Weibo</b>	<b>Blogger</b>	rakuten	<b>Wikipedia</b>	Globo	<b>Wikipedia</b>
8	Google	<b>Twitter</b>	<b>livdoor</b>	<b>LinkedIn</b>	<b>Orkut.com.br</b>	Windows Live
9	ifeng	eBay	<b>Facebook</b>	<b>Twitter</b>	Yahoo!	<b>Twitter</b>
10	Yahoo	Craigslist	<b>Wikipedia</b>	Rediff	<b>Orkut.com</b>	<b>QQ</b>





Alexa as of May 2009	China	USA	Japan	India	Brazil	Global
1	Baidu	Google	Yahoo.jp	Google.in	Google	Google
2	<b>QQ</b>	Yahoo!	<b>FC2</b>	Google	<b>Orkut.br</b>	Yahoo!
3	Sina	<b>Facebook</b>	Google.jp	Yahoo	Windows Live	<b>YouTube</b>
4	Google.cn	<b>YouTube</b>	<b>YouTube</b>	<b>Orkut.in</b>	Universo Online	<b>Facebook</b>
5	Taobao	<b>Myspace</b>	Rakuten	<b>YouTube</b>	<b>YouTube</b>	Windows Live
6	163	MSN	Livedoor	<b>Blogger</b>	Globo	MSN
7	Google	Windows Live	<b>Ameblo.jp</b>	Rediff	MSN	<b>Wikipedi a</b>
8	Sohu	<b>Wikipedia</b>	<b>mixi</b>	<b>Facebook</b>	Google	<b>Blogger</b>
9	Youku	Craigslist	<b>Wikipedi a</b>	<b>Wikipedi a</b>	Yahoo!	Baidu
10	Yahoo	EBay	Google	Windows Live	Terra	<b>Myspace</b>

Alexa as of August 2011	China	USA	Japan	India	Brazil	Global
1	Baidu	Google	Yahoo.jp	Google.in	Google.br	Google
2	<b>QQ</b>	<b>Facebook</b>	Google.jp	Google	Google	<b>Facebook</b>
3	Sina	Yahoo!	<b>FC2</b>	<b>Facebook</b>	<b>Facebook</b>	<b>YouTube</b>
4	Taobao	<b>YouTube</b>	<b>YouTube</b>	<b>YouTube</b>	<b>YouTube</b>	Yahoo!
5	Google.hk	Amazon	Google	Yahoo!	Universo Online	<b>Blogger</b>
6	163	<b>Wikipedia</b>	<b>Ameblo.jp</b>	<b>Blogger</b>	Windows Live	Baidu
7	<b>Weibo</b>	<b>Blogger</b>	rakuten	<b>Wikipedi a</b>	Globo	<b>Wikipedi a</b>
8	Google	<b>Twitter</b>	<b>livedoor</b>	<b>LinkedIn</b>	<b>Orkut.co m.br</b>	Windows Live
9	ifeng	eBay	<b>Facebook</b>	<b>Twitter</b>	Yahoo!	<b>Twitter</b>
10	Yahoo	Craigslist	<b>Wikipedi a</b>	Rediff	<b>Orkut.co m</b>	<b>QQ</b>



# The Brave New Words

博客

維基

AVATAR

头像

tag cloud

推特

unfriend

tweet

blogosphere

twitterati

defriend

SEXTING

hashtags

Folksonomy

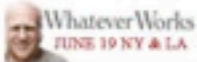


# Politics

HOME PAGE TODAY'S PAPER VIDEO MOST POPULAR TIMES TOPICS

**The New York Times**  
Friday, June 19, 2009

**News**

Search All NYTimes.com   

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION ARTS STYLE TRAVEL JOBS REAL ESTATE AUTOS


**The Lede**

[The New York Times News Blog](#)

June 2, 2009, 7:05 PM

## China's Great Firewall Blocks Twitter

By ROBERT MACKEY



Catherine Henriette/Agence France-Presse — Getty Images

Search This Blog

Previous Post: [Bloggers Ponder Last Message From Missing Jet's Computer](#)

Next Post: [Punditry From Bin Laden and Zawahiri on Obama's Trip to the Middle East](#)

**Recent Posts**

June 18 (38 comments) [Latest Updates on Iran's Disputed Election](#)  
To supplement reporting from New York Times correspondents inside Iran on Thursday, The Lede will continue to track the aftermath of Iran's disputed presidential election online.

June 17 (129 comments) [Wednesday: Latest Updates on Iran's Disputed Election](#)  
On Wednesday, The Lede will continue to track the aftermath of Iran's disputed presidential election online, to supplement reporting from New York Times correspondents inside Iran.

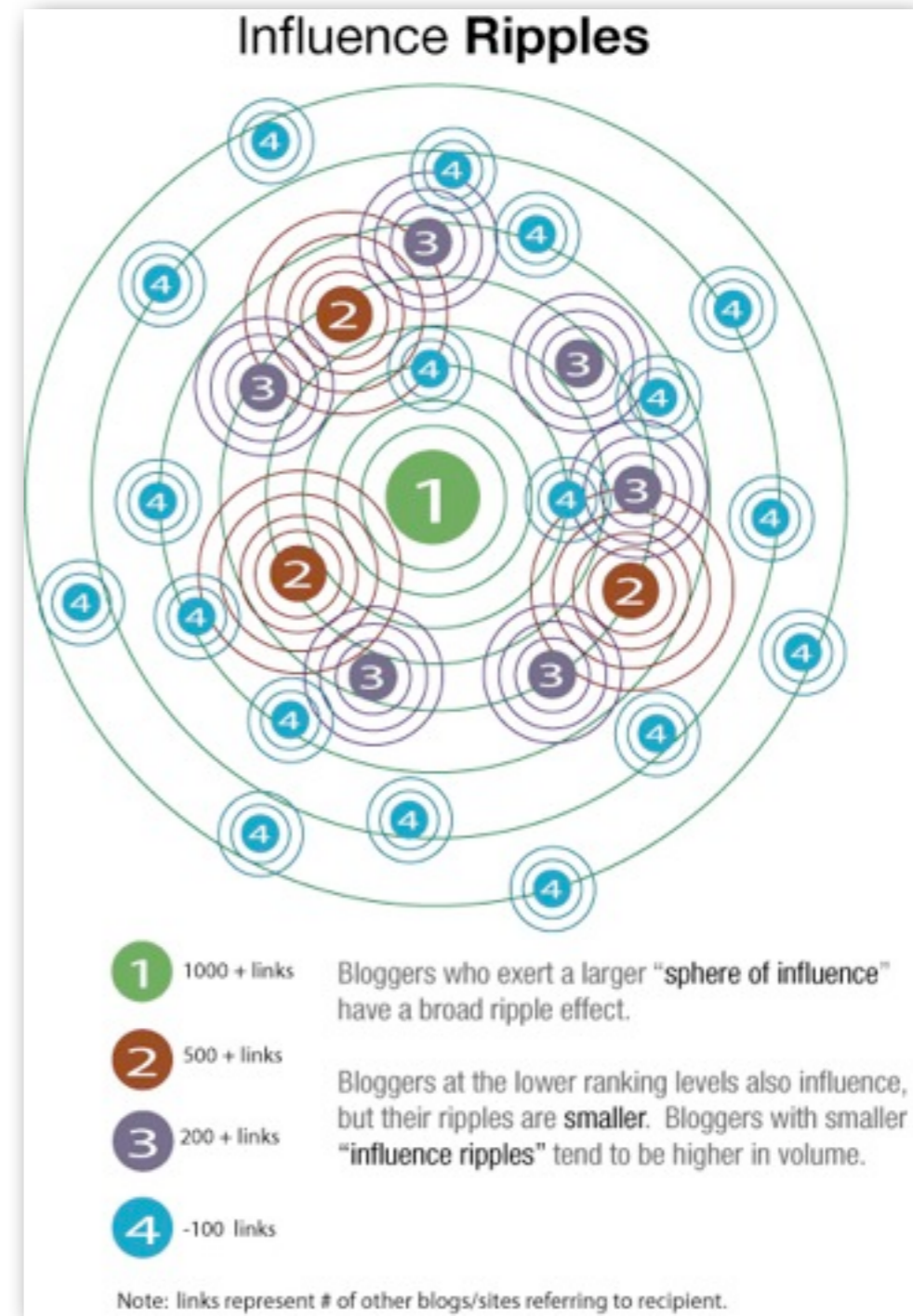
June 16 (198 comments) [Tuesday: Latest Updates on Iran's Disputed Election](#)  
To supplement reporting from New York Times correspondents inside Iran, The Lede





# Commerce

- Social marketing
- Who are the **brokers**?
- Who can exert the **most influence** on buying/selling?
- How **much** should one advertise?





# Public Health

- People's **behavior** can be monitored
- What is on people's mind translates to **search queries**
- Google predicts flu trends...

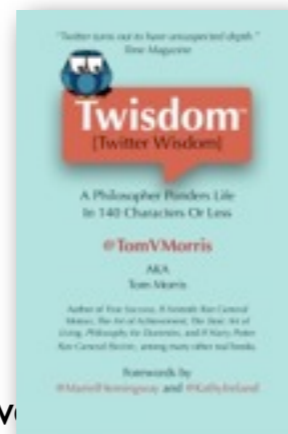
2007–2008 U.S. Flu Activity - Mid-Atlantic Region

ILI percentage



# Pop Culture

- Twisdom: Twitter Wisdom
  - A Philosopher Ponders Life in 140 Characters or Less
    - “I don’t know the key to success, but the key to failure is trying to please everybody.” Bill Cosby Do what you know in your soul is right!
    - It is a miserable state of mind to have few things to desire, and many things to fear. – Francis Bacon
- The Longest Poem In the World-the awesome twitter poem! 956,644 verses this morning and ~4,000 a day!





# The Social Media Generation

The screenshot shows the YouTube channel page for 'Oscar®' (Oscars's Channel). The header features the Academy of Motion Picture Arts and Sciences logo and navigation links for 'VISIT OSCARS.ORG', 'BECOME A FAN', and 'SIGN UP FOR NEWS'. The channel name 'Oscar®' is displayed with a 'Subscribe' button and tabs for 'Uploads' and 'Favorites'. A search bar is located in the top right of the video area.

The main video player shows a large, ornate theater stage with a blue and gold color scheme, filled with people. The video title is 'Steve Martin and Alec Baldwin hosting the Oscars®', with 61 ratings and 312 views. The description reads: 'Steve Martin and Alec Baldwin, co-hosts of the 82nd Academy Awards®, in their opening monologue.' Below the video are options for 'Info', 'Comments', 'Favorite', 'Share', 'Playlists', and 'Flag'. A link says 'View comments, related videos, and more'.

On the right side, a list of related videos is shown with the following titles and view counts:

- Opening Number at the 2010 Oscars® (303 views - 4 hours ago)
- "The Hurt Locker" winning Best Picture (303 views - 4 hours ago)
- John Hughes Tribute at the Oscars® (301 views - 5 hours ago)
- Kathryn Bigelow winning the Oscar® for Directing (301 views - 5 hours ago)
- Sandra Bullock winning Best Actress (309 views - 5 hours ago)
- Jeff Bridges winning Best Actor (334 views - 5 hours ago)
- Steve Martin and Alec Baldwin hosting the (312 views - 6 hours ago)
- Editing Oscar® Nominees (27,246 views - 4 days ago)





# The Age of FaceBook

The screenshot shows the Facebook profile of Barack Obama. The page header includes the Facebook logo, a search bar, and navigation links for Home, Profile, and Account. The profile picture is a portrait of Barack Obama. Below the profile picture are links to 'Add to My Page's Favorites' and 'Suggest to Friends'. A bio section states: 'This page is run by Organizing for America, the grassroots organization for President Obama's agenda for change. To visit the White House Facebook page, go to: <http://bit.ly/2bVCm>. OFA is a special project of the Democratic National Committee.'

The main content area shows three posts:

- Post 1:** A post by Barack Obama with the text: 'Barack Obama 8: the number of people every minute who are denied coverage, charged a higher rate, or otherwise discriminated against because of a pre-existing condition.' It includes a graphic with the number '8' and the title 'Health Reform by the Numbers: 8' with a link to [www.whitehouse.gov](http://www.whitehouse.gov). The post is dated '27 minutes ago' and has 4,913 views.
- Post 2:** A post by Barack Obama titled 'Barack Obama Speaking about health insurance reform this morning at Arcadia University - starting at 11:00 a.m. ET.' It includes a 'LIVE' video player and the title 'President Obama Speaks on Health Insurance Reform' with a link to [www.whitehouse.gov](http://www.whitehouse.gov). The post is dated 'Yesterday at 12:21am' and has 12,287 views.
- Post 3:** A post by Barack Obama with the text: 'Barack Obama I need your help in urging all Americans who want health reform to make their voices heard.' It includes a video player and the title 'President Obama's message to supporters: "We need you in this final march for reform"' with a link to [www.youtube.com](http://www.youtube.com). The post is dated 'March 5 at 8:14am' and has 22,867 views.

On the right side of the page, there is a sidebar with a 'Create an Ad' section and a 'Connect With More Friends' section featuring an envelope icon and the text: 'Share the Facebook experience with more of your friends. Use our simple invite tools to start connecting.' Below this is a 'More Ads' link.





# Outline

- Introduction to Social Computing
- Social Network Theory
- Graph mining
- Ranking and Link Analysis
- Recommender Systems
- Human Computation
- Opinion Mining/Sentiment Analysis
- Opinion mining and sentiment analysis
- Social Computing in Education
- Social Monetization
- and possibly more...



# Web 2.0

- Web as a medium vs. **Web as a platform**
- Read-Only Web vs. **Read-and-Write Web**
- Static vs. **Dynamic**
- Restrictive vs. **Freedom & Empowerment**
- Technology-centric vs. **User-centric**
- Limited vs. **Rich User Experience**
- Individualistic vs. **Group/Collective Behavior**
- Consumer vs. **Producer**
- Transactional vs. **Relational**
- Top-down vs. **Bottom-up**
- People-to-Machine vs. **People-to-People**
- Search & browse vs. **Publish & Subscribe**
- Closed application vs. **Service-oriented Services**
- Functionality vs. **Utility**
- Data vs. **Value**



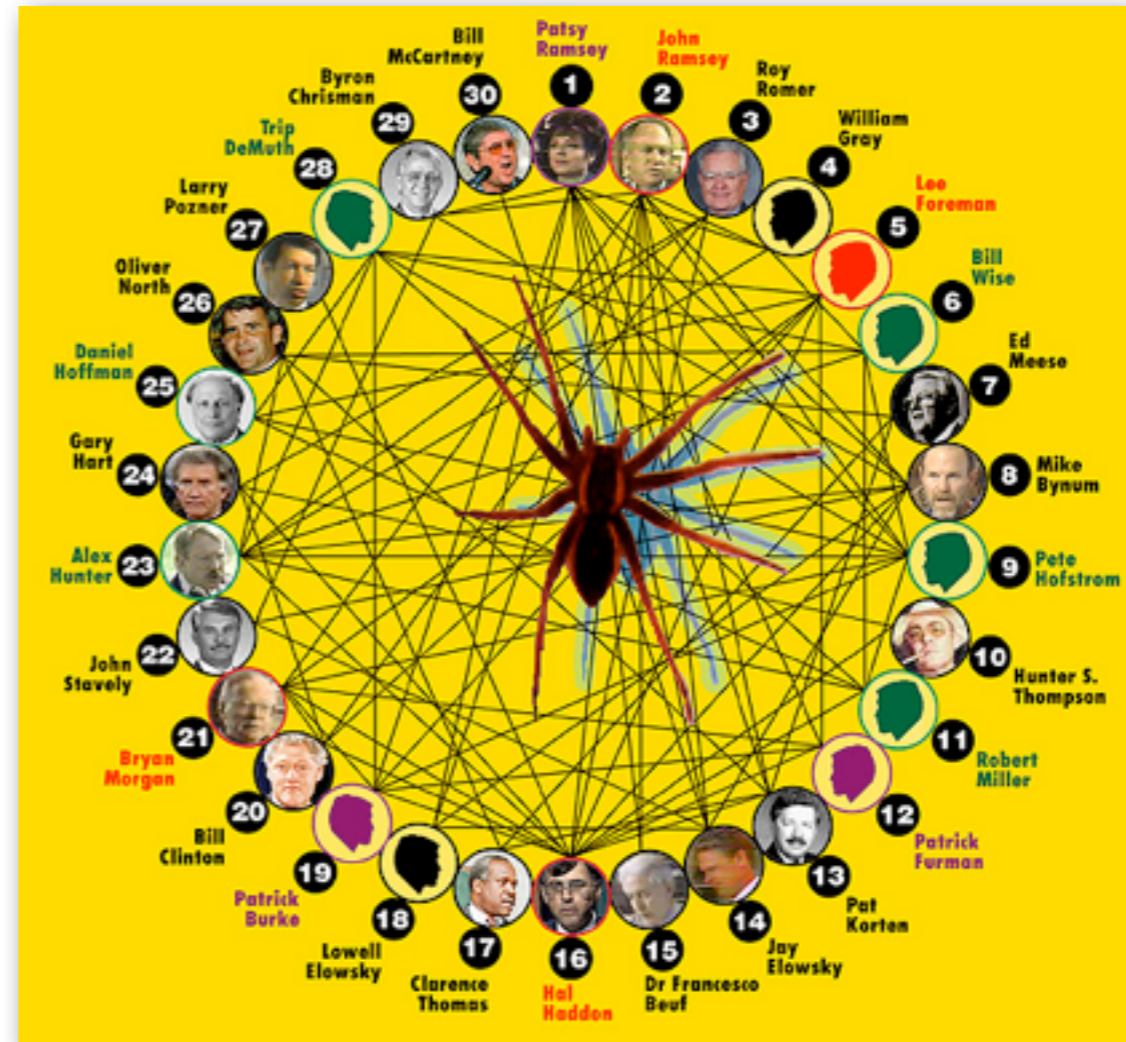


# Social Networks

Society:

**Nodes:** individuals

**Links:** social relationship  
(family/work/friendship/etc.)

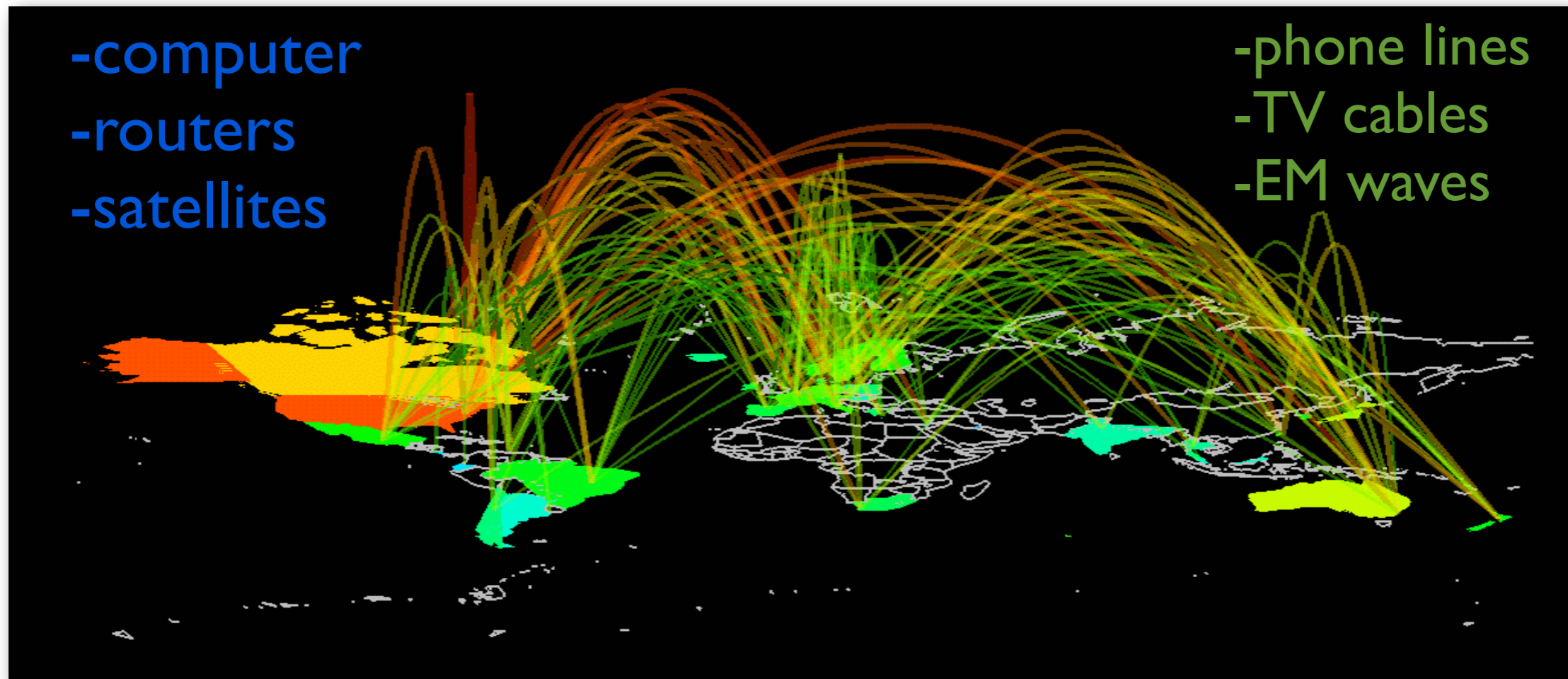


S. Milgram and John Guare: **Six Degree of Separation.**  
Social networks: Many **individuals** with diverse **social interactions** between them.



# Social Networks

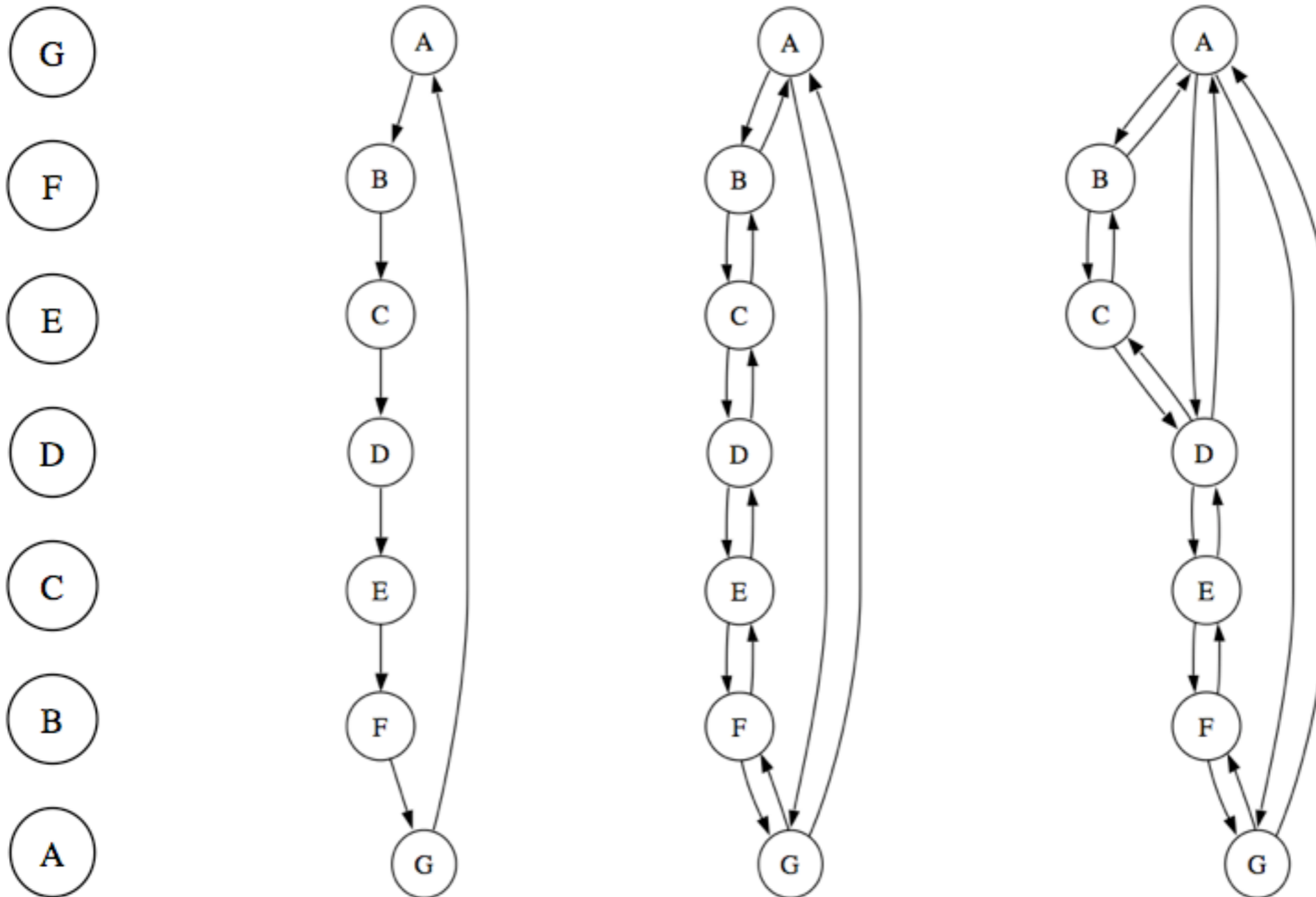
- The Earth is developing an electronic nervous system, a network with diverse **nodes** and **links**.



Communication networks: many non-identical components with diverse connections between them.

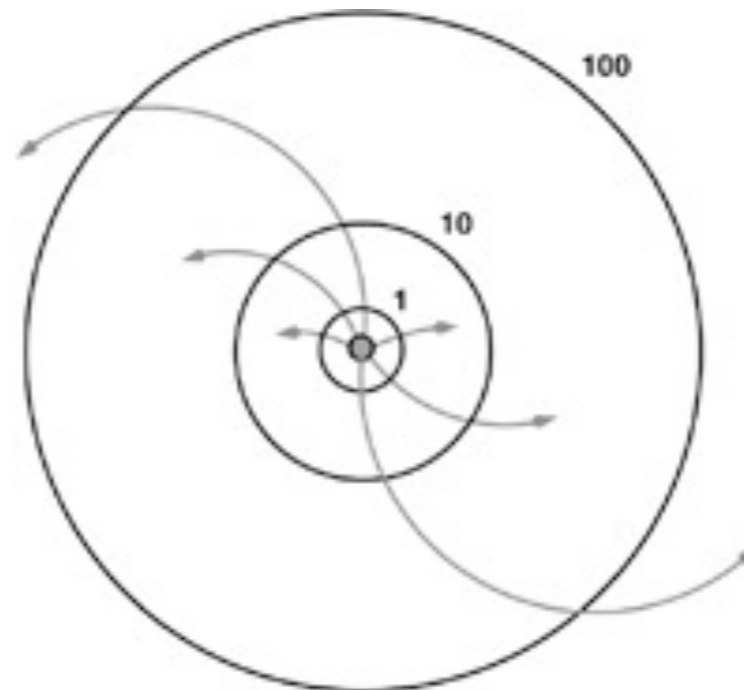
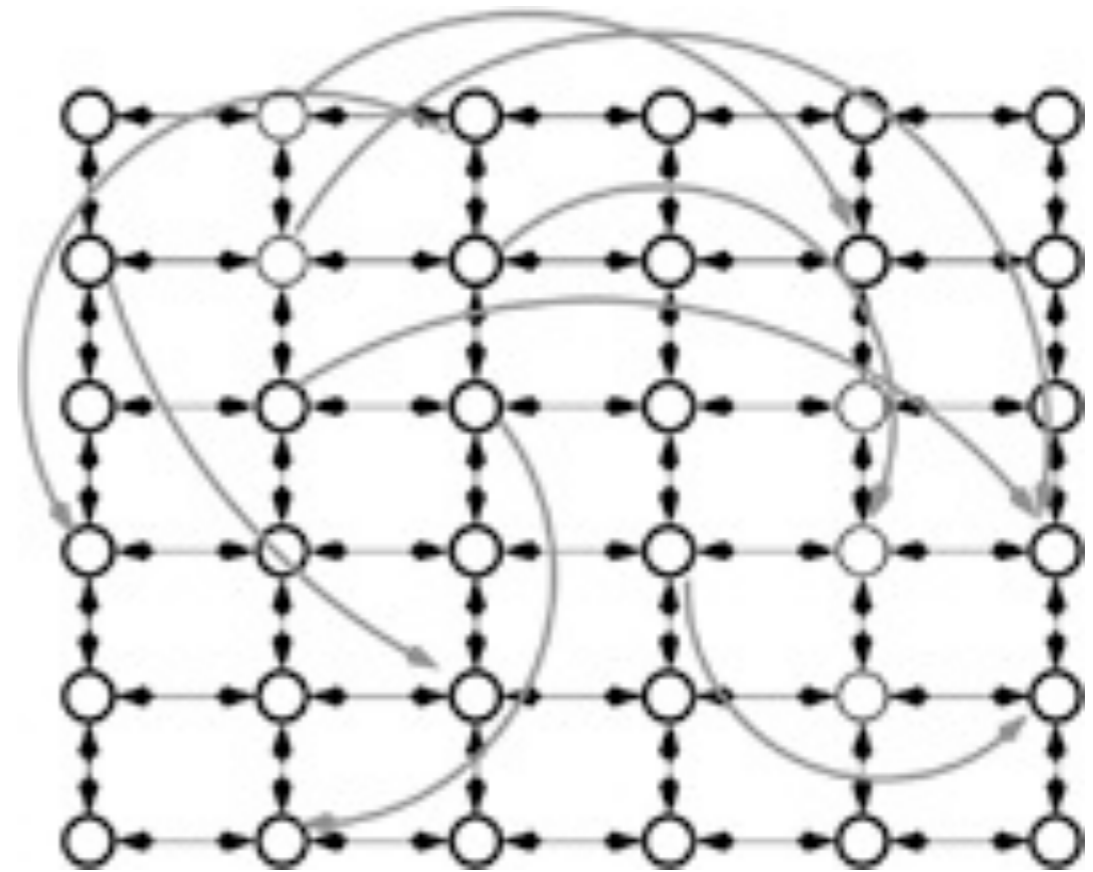
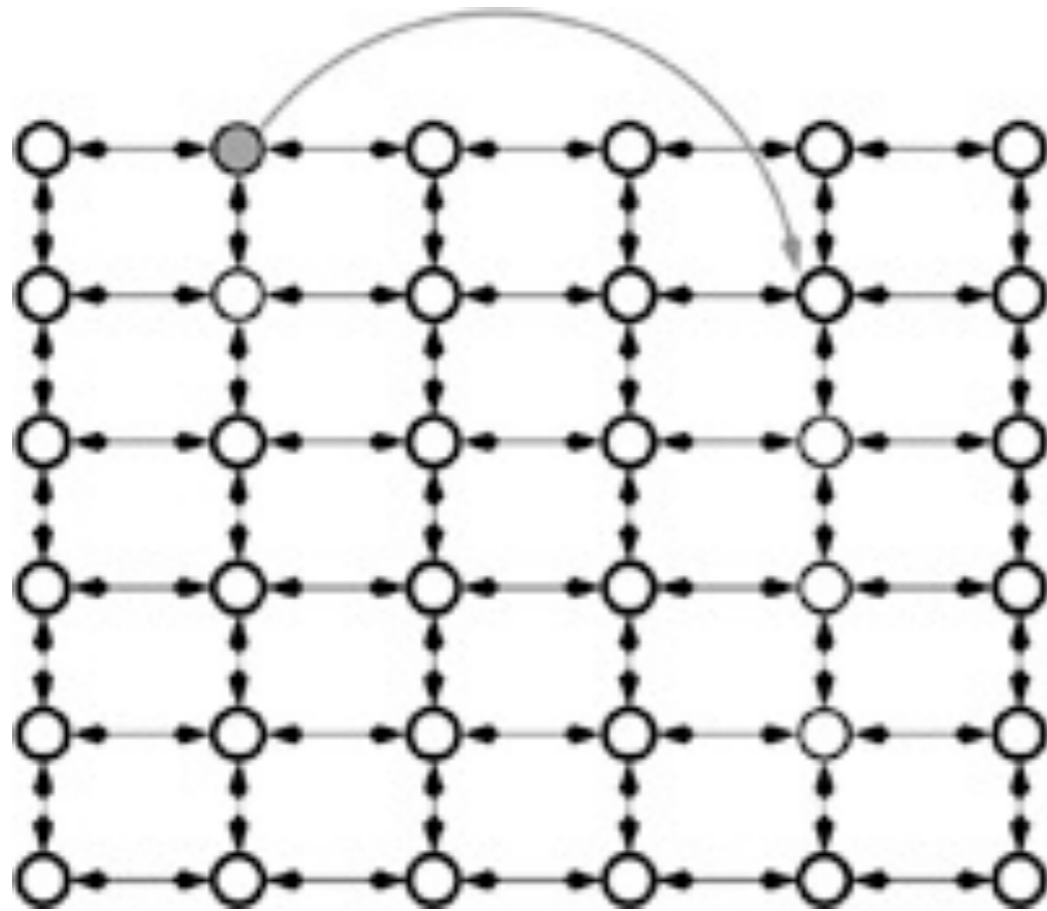


# The Flow of Information





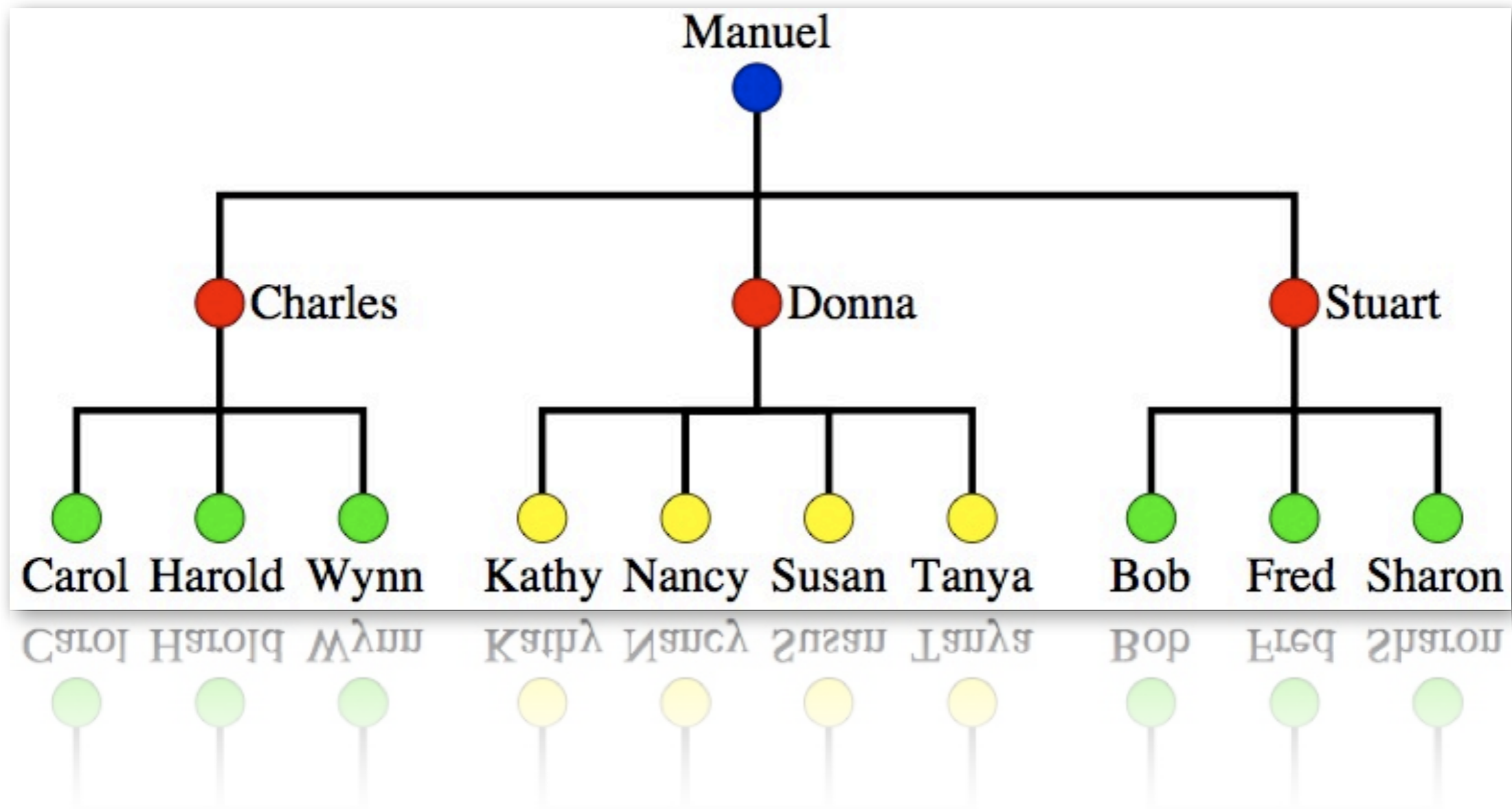
# Examples



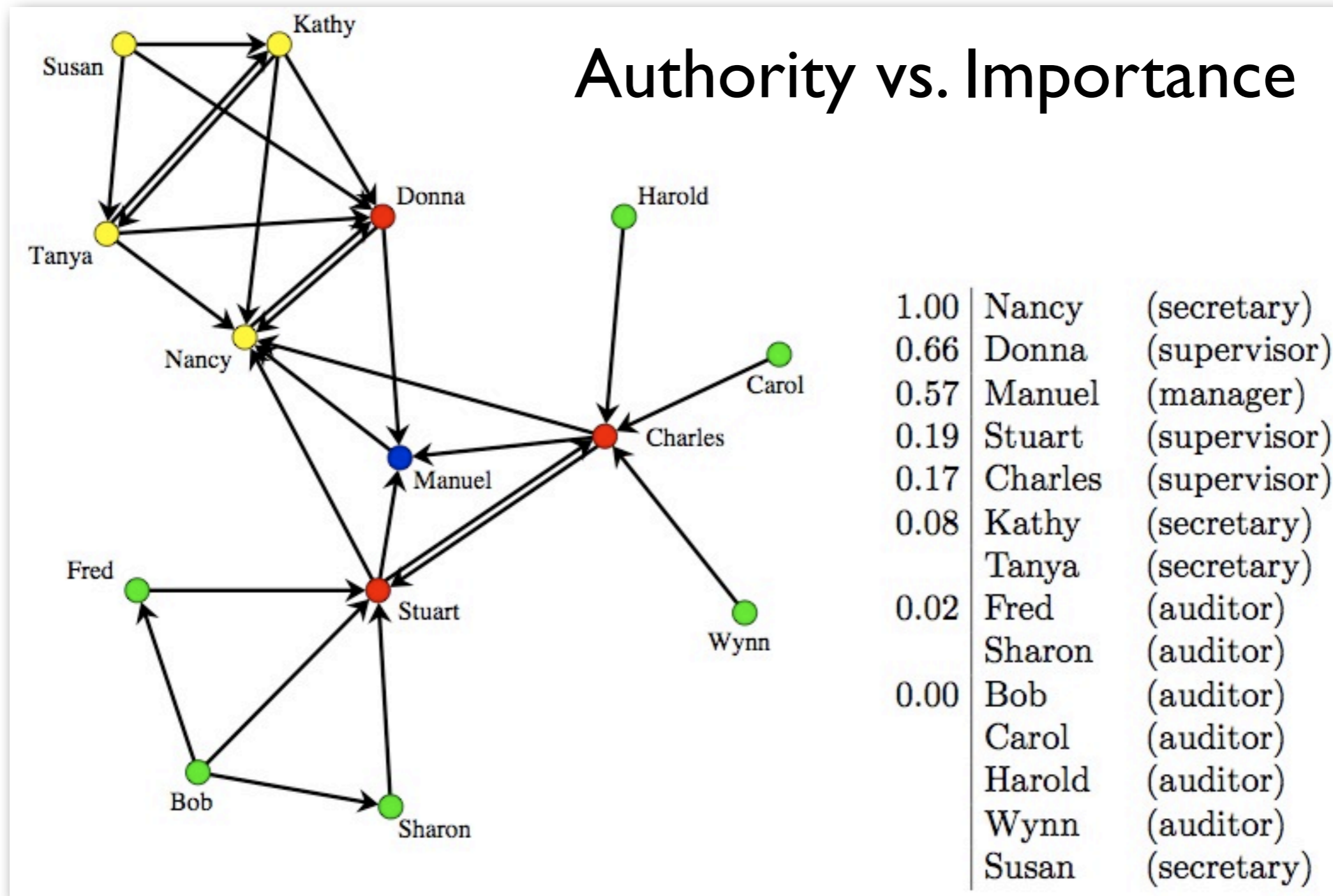
[Kleinberg 1999]



# Organizational Chart



# Social Network Chart





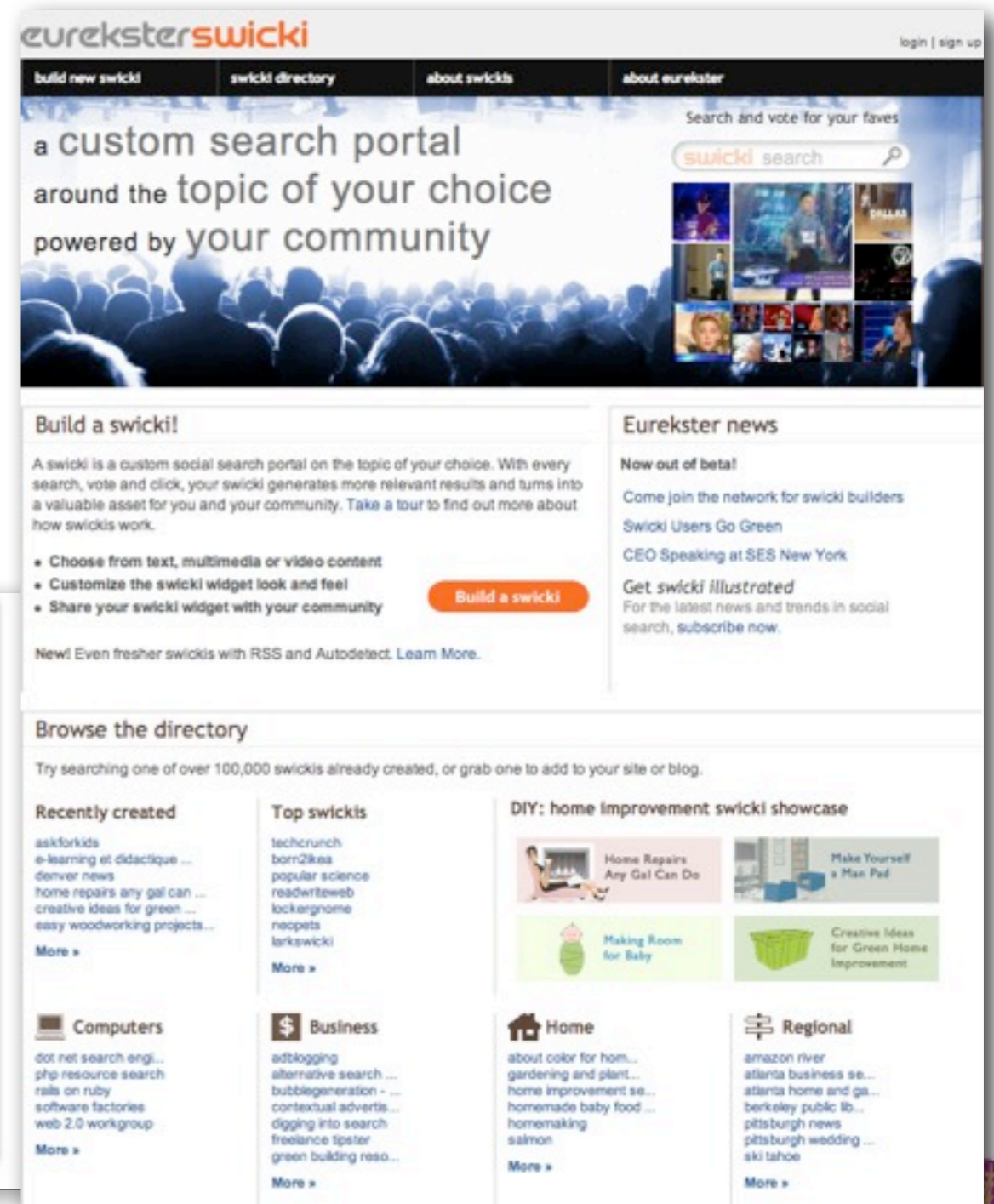
# Social Networking Sites

- Example of Social Networking Sites: FaceBook, MySpace, Blogger, QQ, etc.



# Social Search

- Social Search Engine
- Leveraging your social networks for searching





# Social Media

The screenshot shows the YouTube homepage with the following elements:

- Header:** YouTube logo with the tagline "Broadcast Yourself™". Navigation tabs for Home, Videos, Channels, and Community. A search bar and an "Upload" button.
- Videos being watched right now...:** A row of five video thumbnails with their durations (02:13, 03:29, 01:58, 07:01, 03:53).
- Promoted Videos:** Four video thumbnails with titles like "Think Again Awards" and "第14屆十大電視廣告頒獎典禮 - 飛出...".
- Featured Videos:** A list of featured videos with titles and view counts:
  - David Sedaris delivers a pizza:** From [weeknight](#), Views: 11,313, 5 stars, 01:01. More in [Comedy](#).
  - Erbert and Gerbert's Candle Cannon:** From [candlecannon](#), Views: 109,029, 5 stars, 02:34. More in [Entertainment](#).
  - Girl's Night Out:** From [daniovine](#), Views: 169,435, 5 stars, 03:49. More in [Comedy](#).
  - Lionel Neykov - Freeze My Senses:** From [LionelNeykov](#), Views: 150,758, 5 stars, 03:35. More in [Music](#).
- What's New:** A yellow box containing:
  - YouTube Mobile:** "New! Watch ALL YouTube videos on your mobile device".
  - Warp!** "Visually fly through YouTube videos in the Fullscreen player".
  - RSS Feeds:** "Click on the 'RSS this page' link to get fresh videos delivered".
  - SXSW on YouTube:** "For the next week and a half, the SXSW festival is taking over Austin, Texas, to celebrate music, film and all things interactive. [Read more in our Blog](#)".
- Login:** A form with fields for Username and Password, a "Login" button, and links for "Sign Up | Help", "Forgot Username | Forgot Password", and "Login with your Google account".

The screenshot shows the Flickr homepage with the following elements:

- Header:** Flickr logo and a "Sign In" link.
- Main Content:** A large photo of a small plant growing in a crack in the pavement. Text reads: "Share your photos. Watch the world." Below this is a search bar and a "SEARCH" button.
- Statistics:** "3,802 photos uploaded in the last minute · 558,832 photos tagged with urban · 2.2 million photos uploaded this month · [Take the tour](#)".
- Navigation:** Four icons with labels: "Share & stay in touch", "Upload & organize", "Make stuff!", and "Explore...".
- Footer:** A "Take the Tour" button and a link to "Explore Flickr Blog, the World Map, Camera Finder or interesting photos from the last 7 days".

The screenshot shows the Second Life homepage with the following elements:

- Header:** Second Life logo with the tagline "Your World. Your Imagination." and a "Resident, Login | Join" link.
- Navigation:** Links for "What is Second Life?", "SHOWCASE", "COMMUNITY", "BLOG", and "SUPPORT". A search bar labeled "Search Second Life".
- Main Content:** A large image of a man and a woman flying through the sky. Text reads: "Get Started! Membership is FREE! Second Life is an online, 3D virtual world imagined and created entirely by its Residents." Below this is a button: "Discover a whole new world of friends, fashion, music, videos and fun! Explore the best of Second Life >>".
- Footer:** A section titled "Your Organization in Second Life!" with text: "Find out why your business, school or nonprofit organization should get its own virtual world presence." and a "Visit Second Life Now" button.



# Social News/Mash Up

The Digg website interface features a navigation bar with 'Join Digg', 'About', and 'Login'. Below the navigation bar are tabs for 'All', 'News', 'Videos', 'Images', 'Podcasts', and 'Customize'. A secondary navigation bar lists categories like 'Technology', 'World & Business', 'Science', 'Gaming', 'Lifestyle', 'Entertainment', 'Sports', and 'Offbeat'. The main content area is titled 'News, Videos, Images' and displays a list of news items with titles, brief descriptions, and user interaction options like 'Share' and 'Bury'. A sidebar on the right is titled 'Top in All Topics' and lists trending articles.

The Twitter website features a blue header with the 'twitter' logo and a 'Select Language' dropdown. The main content area is titled 'What is Twitter?' and includes a 'Watch a video!' button. Below the title is a large illustration of a yellow bird on a branch. To the right is a sign-in form with fields for 'user name or email address' and 'password', along with a 'Remember me' checkbox and a 'Sign in' button. Below the sign-in form is a link for 'Forgot password? Click here.' and a green button for 'Already using Twitter from your phone? Click here.'

The FoxyTunes website features a search bar and navigation tabs for 'Albums' and 'Tracks'. The main content area is dedicated to the artist Björk, showing a video player for 'All is full of love' and a list of lyrics from Yahoo! Music. The lyrics list includes '5 Years', 'Alarm Call', 'All Is Full of Love', 'All Neon Like', 'An Echo, A Stain', 'Army of Me', 'Aurora', 'Bachelorette', 'Big Time Sensuality', 'Cetacea', 'Cocoon', 'Come to Me', 'Crying', and 'Desired Constellation'. There is also a 'Flickr Photos' section and a 'Music on Hype Machine' section.

The TwitterVision website features a map of the United States with a 'twittervision' logo in the top right corner. A tweet overlay is shown on the map, indicating a tweet by 'Killane' with the text 'I feel odd' posted '17 minutes ago in North of Seattle'. The map shows various geographical features like the Chukchi Sea, Beaufort Sea, Gulf of Alaska, and North Pacific Ocean.





# Social News and Knowledge Sharing

- Social news refers to websites where users can submit their own information. Users can also vote on news or other links to determine which links are presented

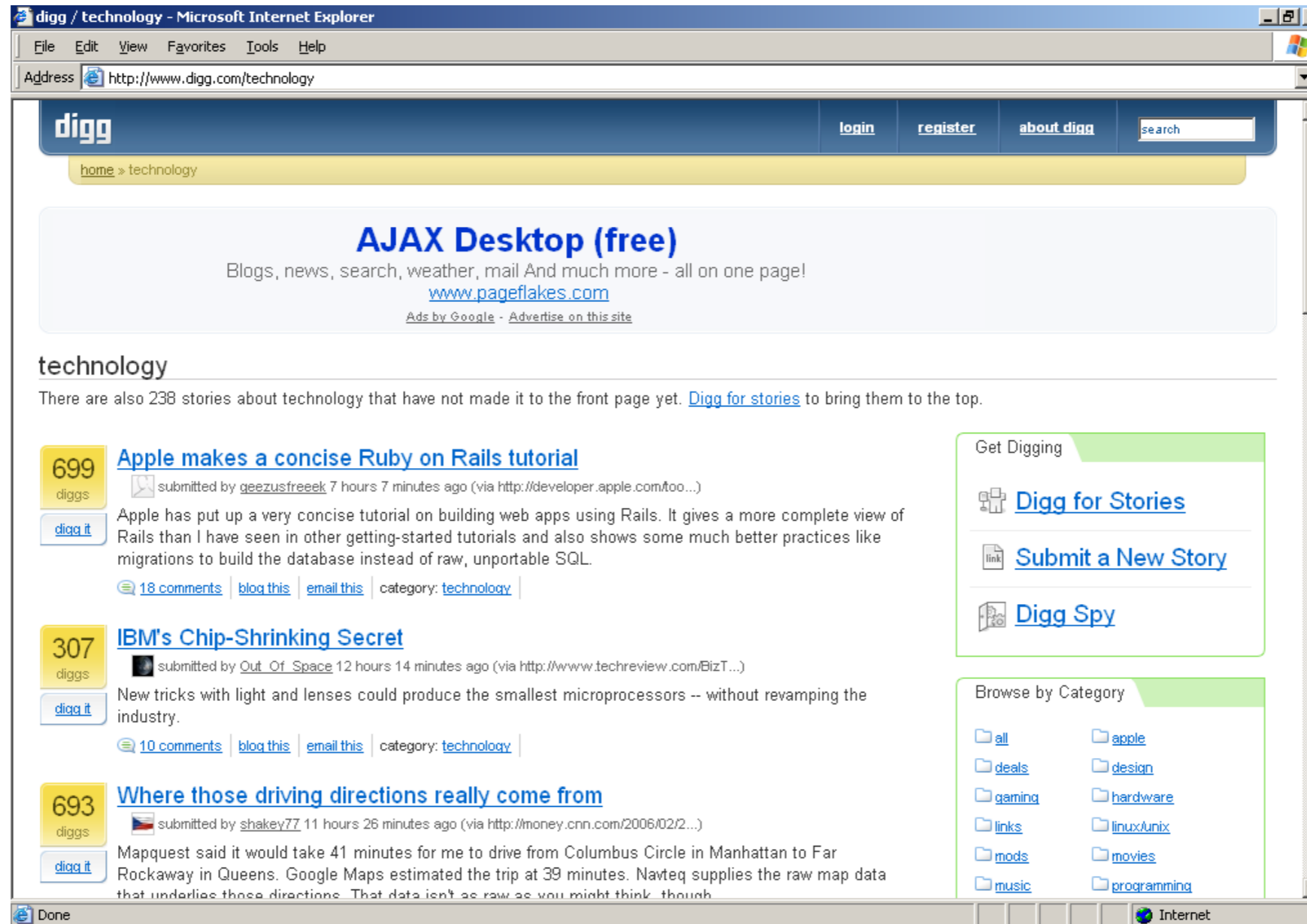


Wikis allow collaborative publishing.  
Anyone can create new articles or edit existing articles.  
All versions are kept.





# Social News and Knowledge Sharing



- **Digg.com** members “vote” for stories to appear on the home page



# Social News and Knowledge Sharing

- The notion that each individual contributes to a collective pool of knowledge is further realized in AnswerBus, Webclopedia, Yahoo's babelfish, etc.

**AnswerBus**

Ask

Type in your question in English, French, Spanish, German, Italian or Portuguese.

**Webclopedia**

*Targeted Delivery of Multilingual Information*

**YAHOO!** BABEL FISH



# Social News and Knowledge Sharing

- Question and answering

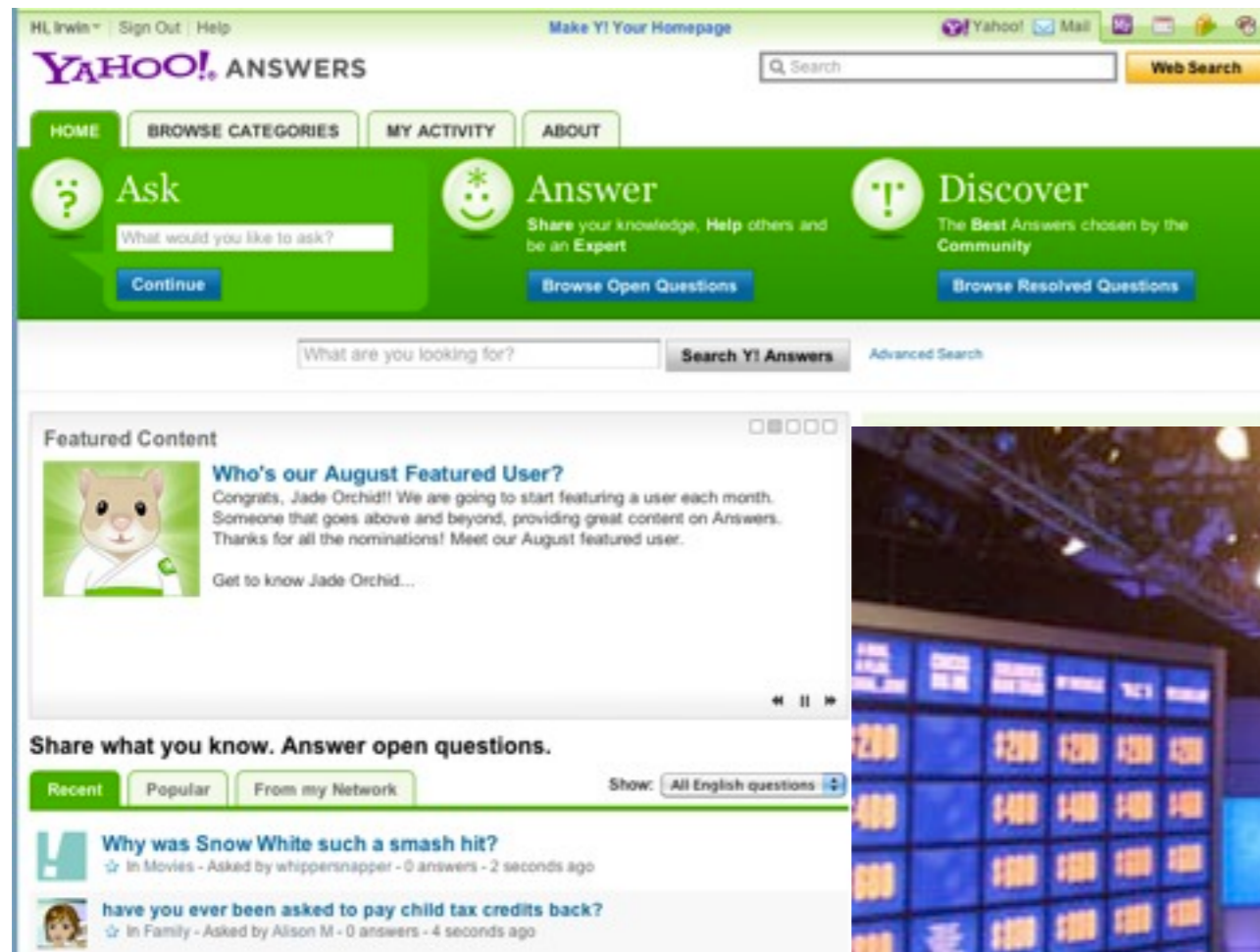


PHOTO: JULIANNE PEPITONE/CNNMONEY





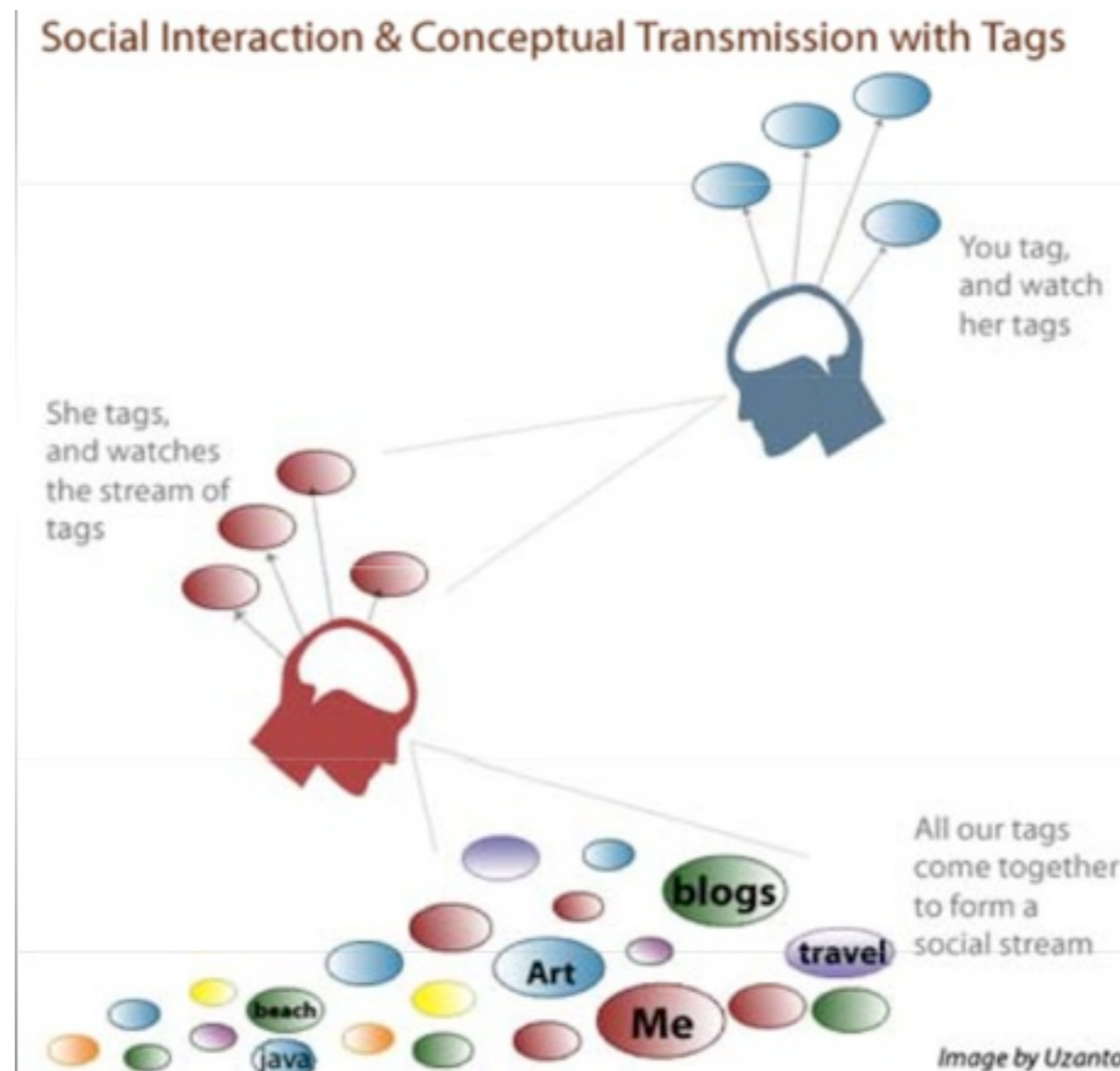
# Social Bookmarking

- **What is a tag?**
  - Descriptive metadata
  - A keyword or term associated with or assigned to a piece of information
  - User defined, created and shared
  - Many web users do it every day, with very little conscious awareness that they are “cataloging”
- **What gets tagged?**
  - Pictures, blog posts, video clips, catalog entries, just about anything...



# Social Bookmarking

- Share one's tags
- Make the individual browsing experience a social one



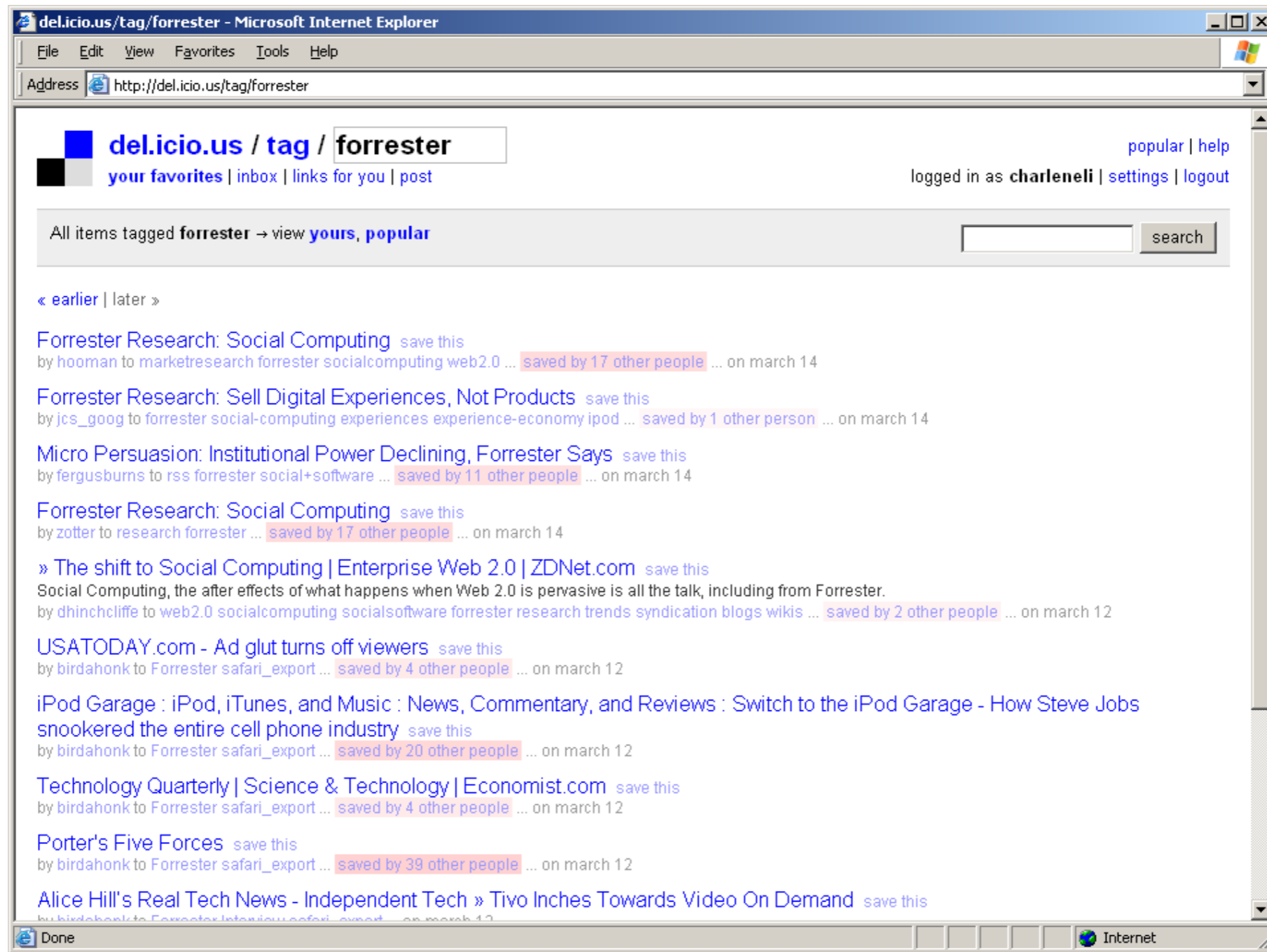
# Why users tag?

- Tagging means something specific to the user
- It is easy -- anyone can do it
- Finding things on the Internet
- Serendipitous discovery
- It is social
- New ways to share and discover





# Social Bookmarking in del.icio.us



# Social Bookmarking in StumbleUpon

- **StumbleUpon** allows users to discover and rate web pages, photos, and videos. It chooses which web page to display based on the user's ratings of previous pages, ratings by his/her friends, and by the ratings of users with similar interests.









# Folksonomies

- Folksonomies are the actual output result of collaborative tagging
- Literally, it is taxonomy by “folks”
- Grass-roots
- Community based
- Inclusive -- everyone can get involved
- Scalability



# Tag Clouds

- Visualization of tags
- Weighted value -- based on size, frequency of use of tag



# Social/Human Computation

Security Check: Enter both words below, separated by a space. What's This?  
Can't read this? Try another.  
[Try an audio captcha](#)



Text in the box:

I have read and agree to the [Terms of Use and Privacy Policy](#)

[Sign Up](#)

[Problems signing up? Check out our help pages](#)

Security Check: Enter both words below, separated by a space. What's This?  
Can't read this? Try another.  
[Try an audio captcha](#)



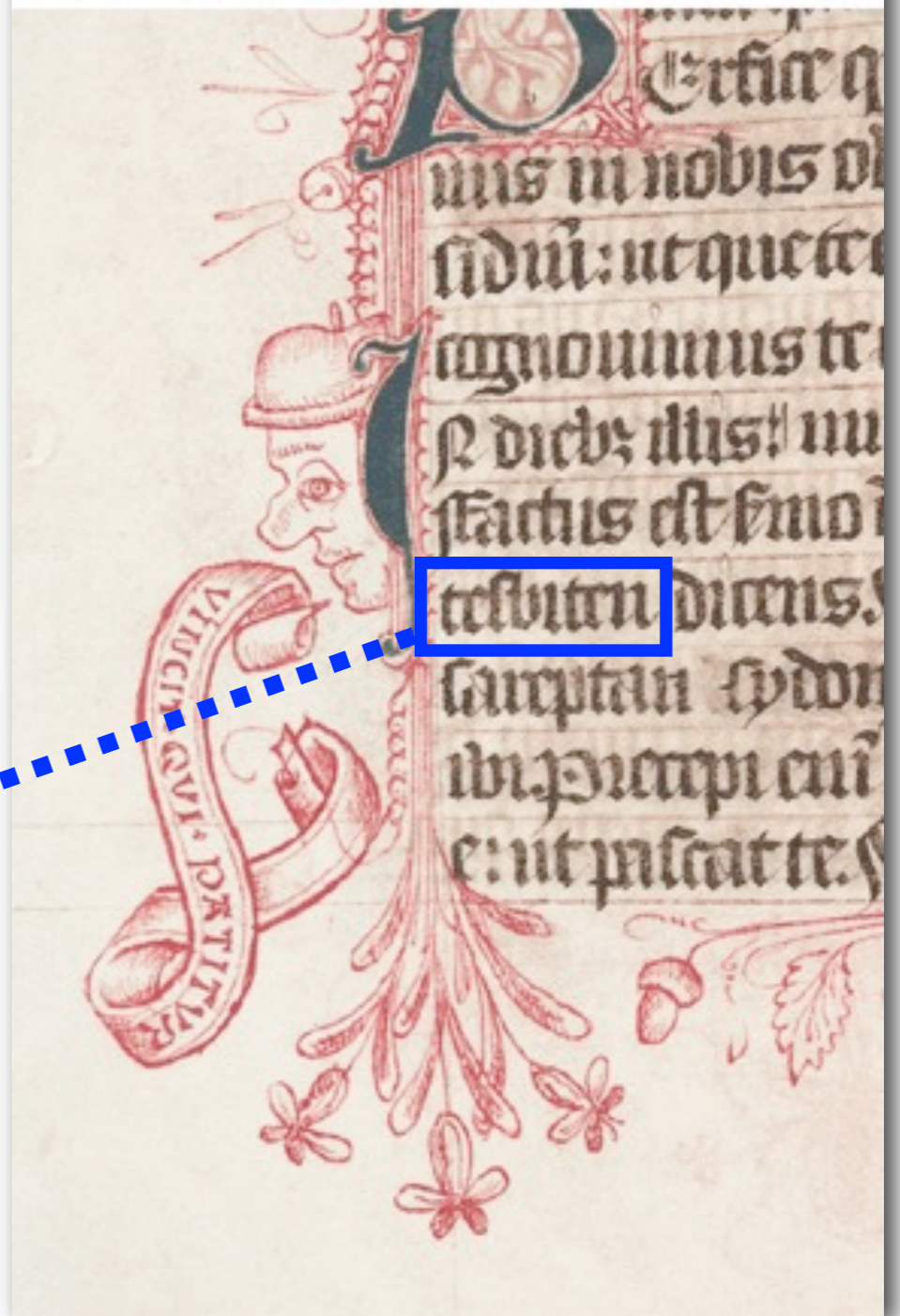
Text in the box:

I have read and agree to the [Terms of Use and Privacy Policy](#)

[Sign Up](#)

[Problems signing up? Check out our help pages](#)

MS. Don. b. 6, fol. 48v (detail) © Bodleian Library, University of Oxford



[Problems signing up? Check out our help pages](#)

[Sign Up](#)





# Human Computation

The screenshot shows the Google Image Labeler interface. At the top left is the Google logo and 'Image Labeler BETA'. The main title is 'Google Image Labeler'. On the left, there is a scorecard with 'time left' at 01:17, 'score' at 0, and 'passes' at 0. In the center, there is a text input field with a 'label' button and a 'pass' button. Below the input field, it says 'Your partner has suggested 10 labels.' A large image of a lake and mountains is displayed. Below the image is a 'zoom out' button. On the right, there is a list of labels: 'off-limits', 'sky', 'water', 'blue', 'lake', 'mountain'. Below the labels is a 'my labels' section. At the bottom, there are links for 'Privacy Policy', 'Terms of Use', and 'Return to Google Image Search', along with a copyright notice '© 2007 Google'. Red starburst shapes are overlaid on the interface, highlighting the input field, the scorecard, the labels list, and the 'my labels' section.



# Crowdsourcing

**amazonmechanical turk**  
Artificial Intelligence

Already have an account?  
Sign in as a Worker | Requester

Your Account | **HITS** | Qualifications

Introduction | Dashboard | Status | Account Settings

**Mechanical Turk is a marketplace for work.**  
We give businesses and developers access to an on-demand, scalable workforce.  
Workers select from thousands of tasks and work whenever it's convenient.  
**210,857 HITS** available. [View them now.](#)

## Make Money by working on HITS

HITS - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITS now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

## Get Results from Mechanical Turk Workers

Ask workers to complete HITS - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITS completed in minutes
- Pay only when you're satisfied with the results



[FAQ](#) | [Contact Us](#) | [Careers at Amazon](#) | [Developers](#) | [Press](#) | [Policies](#) | [Blog](#)  
©2005-2011 Amazon.com, Inc. or its Affiliates

An **amazon.com** company

CALL US: 415 - 651 - 4485 [LOGIN](#)

**CrowdFlower** SOLUTIONS SELF-SERVICE NEWS & EVENTS BLOG

## ENTERPRISE CROWDSOURCING SOLUTIONS

Tailored crowdsourcing solutions to fit your needs.

**CUSTOM SOLUTIONS** [LEARN MORE](#)

Content Generation | Business Listing Verification | Search Relevance | Product Categorization | **Custom Solutions**

### CrowdConf 2011

The world's largest crowdsourcing conference  
November 1-2

[GET TICKETS NOW](#)

#### PRESS



**CrowdFlower Broadens Executive Team**  
TMCTnet.com | August 04, 2011



**San Francisco-Based CrowdFlower Names New CEO and CFO**  
Silicon Valley Wire | August 02, 2011



**CrowdConf2011: Get More Out of Your Crowdsourcing Efforts**

#### EVENTS

**PopTech**  
October 18-22 [See more](#)

**Wolfram Data Summit 2011**  
September 8-9 [See more](#)

**CrowdConf 2011**  
November 1-2 [See more](#)

#### BLOG



**Should organizations establish a Crowdsourcing**



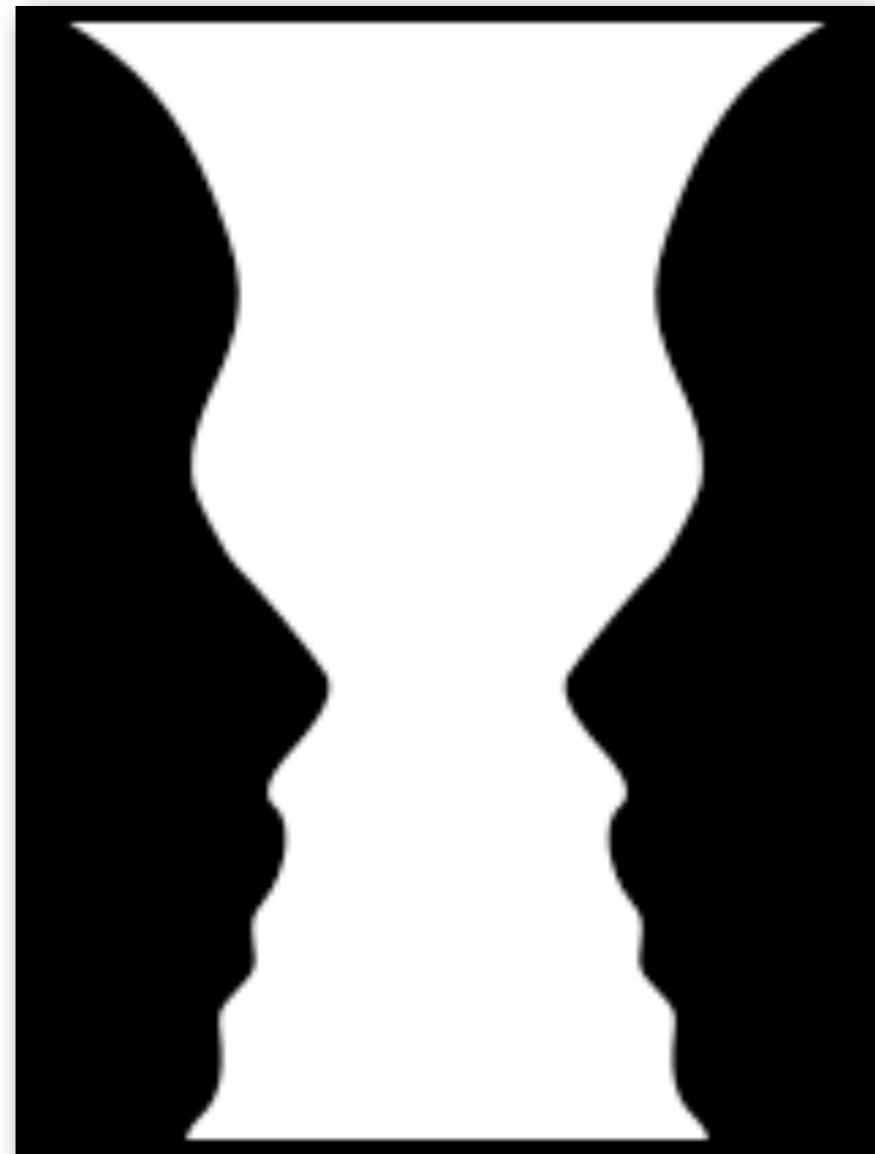
# Web 2.0 Revolution

- **Glocalization**-think globally and act locally!
- **Weblication**-Web is the application!
- Three C's

**C**onnectivity

**C**ollaboration

**C**ommunities



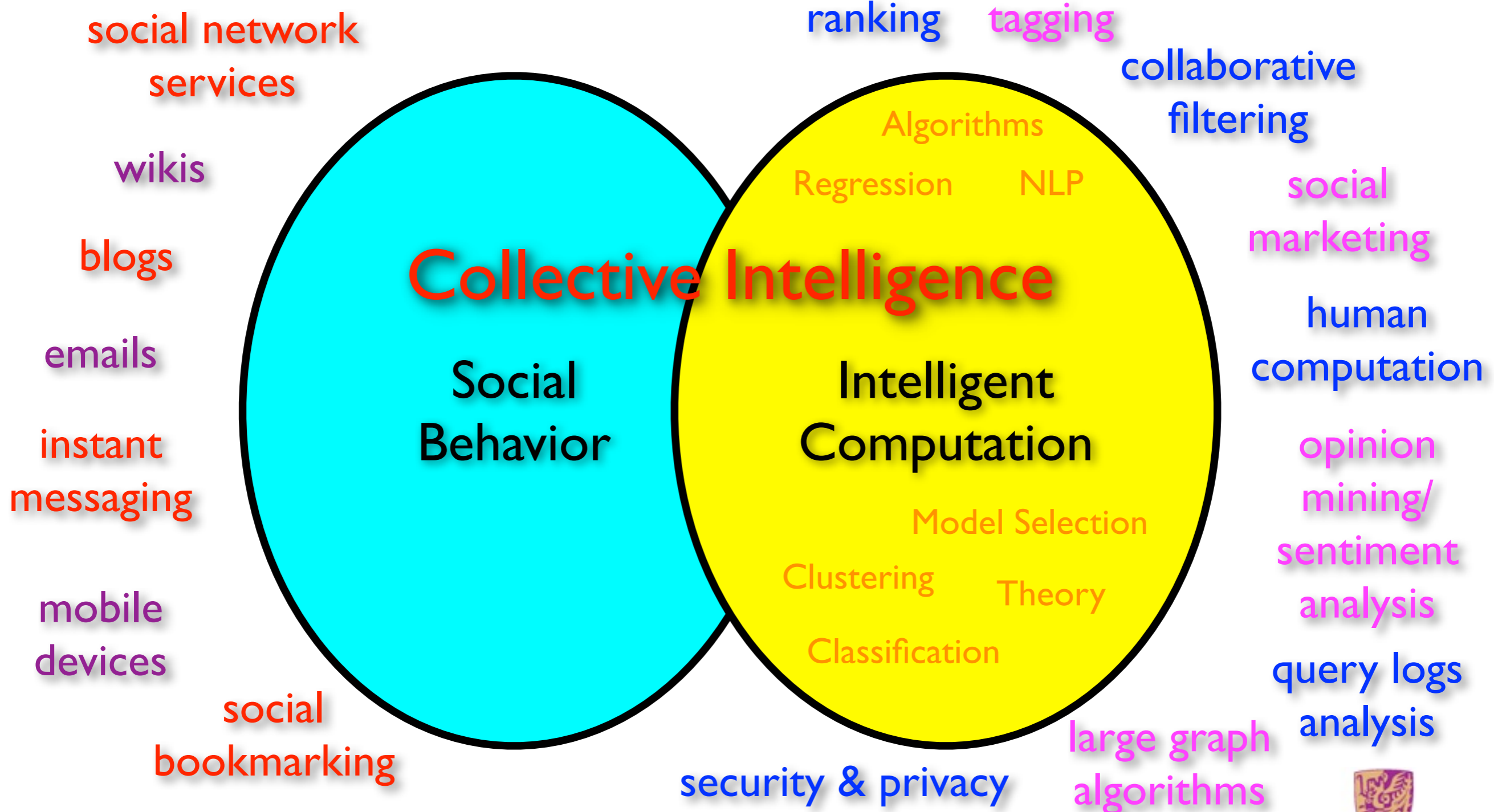


# Social Relations

presence  
identity  
social role  
reputation  
expertise  
trust  
ownership  
accountability  
knowledge  
crew  
teams  
populations  
binary  
cardinal  
integer  
real  
squad  
organizations  
cohorts  
markets  
communities  
partners  
groups



# Social Computing



# Definition of Social Computing [wiki]

- Any Computer-mediated communication and interaction
- In the weaker sense: **supporting any sort of social behavior**
  - blogs, email, instant messaging, wiki, social network services, social bookmarking
- In the stronger sense: **supporting “computations” that are carried out by a group of people**
  - collaborative filtering, online auctions, prediction markets, reputation systems, tagging, verification games





# Emerging Issues

- **Theory** and models
- **Search, mining, and ranking** of existing information, e.g., spatial (relations) and temporal (time) domains
- Dealing with **partial** and **incomplete** information, e.g., collaborative filtering, ranking, tagging, etc.
- **Scalability** and algorithmic issues
- **Security** and **privacy** issues
- **Monetization** of social interactions



# Social Network Theory

- Consider many kinds of networks:
  - social, technological, business, economic, content, ...
- These networks tend to share certain informal properties:
  - **large scale**; continual growth
  - **distributed**, organic growth: vertices “decide” who to link to
  - interaction restricted to **links**
  - mixture of **local** and **long-distance** connections
  - **abstract** notions of distance: geographical, content, social,...



# Six Degree of Separation

- “Six degrees of separation between us and everyone else on this planet” [John Guare, 1990]
- What is the probability of two strangers having a mutual friend?
- What is the chain of intermediaries between two strangers?





# Small World Networks

- A network that most nodes can be reached from every other node by a small number of steps
- $L \propto \log N$  , where  $L$  is the steps and  $N$  is the network size
- Examples: road, power grid, online social networks, email network, neural networks, WWW, etc.



# Dunbar's Number

- It is theoretical **cognitive limit** to the number of people with whom one can maintain stable social relationships
- It is assumed to be between 100 to 230 with 150 as the norm in various studies
- Allen curve--the **exponential drop** of frequency of communication as the distance between them increases



# The Tipping Point

- It is “the moment of **critical mass**, the **threshold**, the **boiling point**”
- Three rules
  - The Law of the Few
  - The Stickiness Factor
    - ...the specific content of a message that renders its impact memorable, i.e., Apple’s 1984 Super Bowl commercial
  - The Power of Context
    - ...are sensitive to the conditions and circumstances of the times and places in which they occur





# Pareto Principle

- Also known as the “80-20 Rule” or “The Law of the Vital Few”
- Roughly 80% of the effects come from 20% of the causes
  - In 1906, the observation that 80% of the land in Italy was owned by 20% of the population
  - 80% of your profits come from 20% of your customers
  - 80% of your complaints come from 20% of your customers
  - 80% of your sales are made by 20% of your sales staff
  - 80% of the work will be done by 20% of the participants



# Social Network Theory

- Do these networks share more **quantitative** universals?
- What would these “universals” be?
- How can we make them precise and measure them?
- How can we explain their universality?
- This is the domain of **social network theory**



# Some Interesting Quantities

- **Connected components**

- how many, and how large?

- **Network diameter**

- maximum (worst-case) or average?
- exclude infinite distances? (disconnected components)
- the small-world phenomenon

- **Clustering**

- to what extent that links tend to cluster “locally”?
- what is the balance between local and long-distance connections?
- what roles do the two types of links play?

- **Degree distribution**

- what is the typical degree in the network?
- what is the overall distribution?





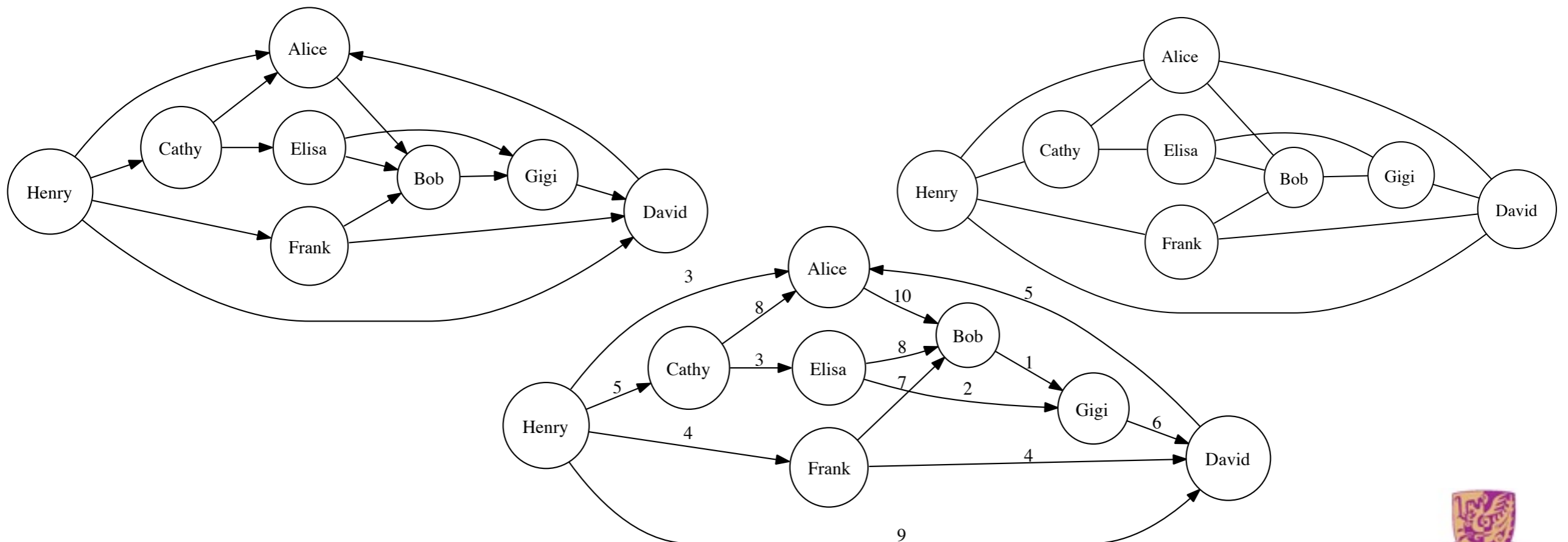
# Types of Relations

- **Kinship**—mother of, wife of
- **Other role-based**—boss of, teacher of, friend of, brother of, father of, sister of, enemy of, lover of
- **Cognitive/perceptual**—knows, aware of what they know, is familiar with
- **Affective**—likes, loves, hates, admires, trusts
- **Interactive**—give advice, talks to, fights with, sex/drugs with, buys from, sells to
- **Affiliations**—belong to same clubs, is physically near
- **Derived**—has subscription to the same magazine as, is taller than, distance between
- **Flows**—moves to, flows to



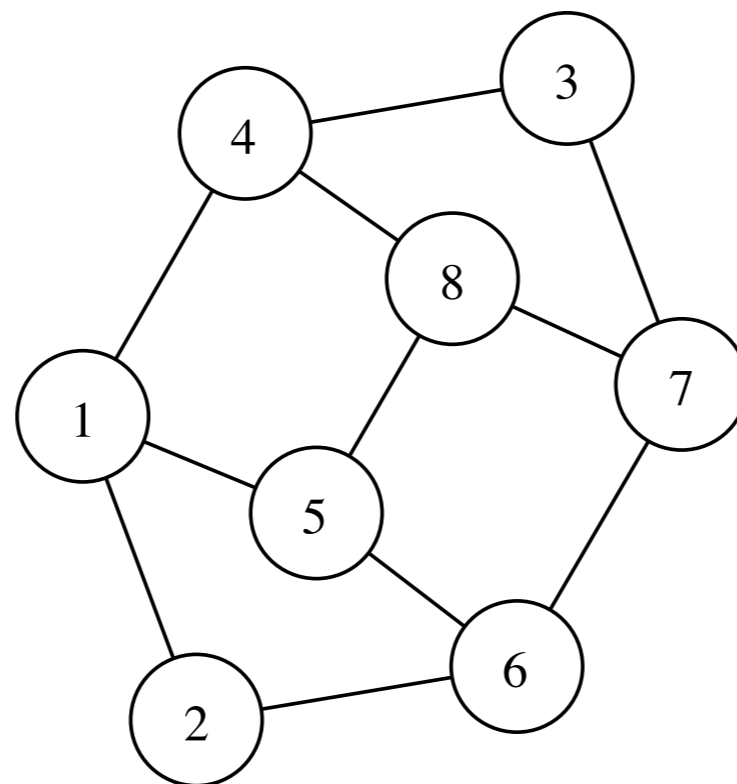
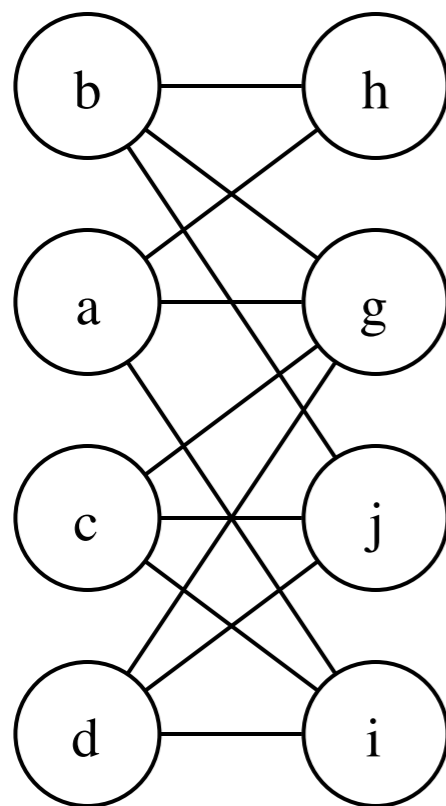
# Define Graphs

- A graph  $G = (V, E)$  consists of a set of vertices,  $V$ , and a set of edges,  $E$ .
- Each edge is a pair  $(v, w)$ , where  $v, w \in V$ . It is said to join the vertices  $v$  and  $w$ .
- If the edge  $e = (v, w) \in E$ , then  $u$  and  $v$  are both said to be *incident* with  $e$  and *adjacent* to each other.
- If the pair is ordered, then the graph is directed (digraphs).
- One can associate an attribute to the edge which is called *weight*.



# Define Graph Isomorphism

- Two graphs  $G$  and  $H$  are said to be *isomorphic*, denoted by  $G \sim H$ , if there is a one-to-one correspondence, called an *isomorphism*, between the vertices of the graph such that two vertices are adjacent in  $G$  if and only if their corresponding vertices are adjacent in  $H$ .
- Likewise, a graph  $G$  is said to be *homomorphic* to a graph  $H$  if there is a mapping, called a homomorphism, from  $V(G)$  to  $V(H)$  such that if two vertices are adjacent in  $G$  then their corresponding vertices are adjacent in  $H$ .



a	1
b	6
c	8
d	3
g	5
h	2
i	4
j	7





# Define Adjacency Matrix

A graph  $G$  with  $n$  nodes can be represented by an  $n$ -by- $n$  matrix. Given  $V = \{v_1, v_2, \dots, v_n\}$ . Then the adjacency matrix  $A$  is an  $n$ -by- $n$  matrix whose entry  $A_{ij}$  is defined to be:

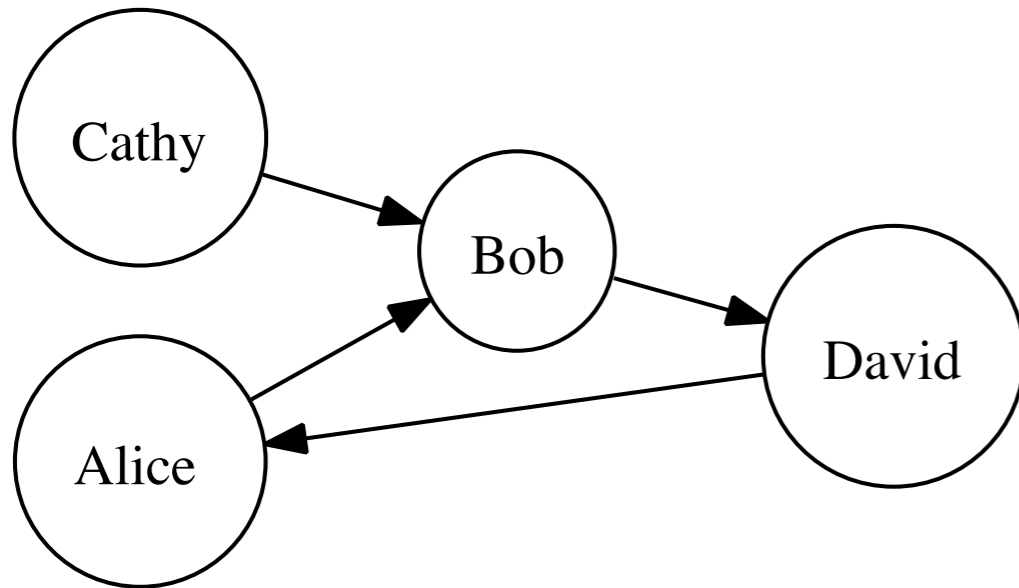
$$A_{ij} = \begin{cases} 1, & \text{if there is an edge from } v_i \text{ to } v_j \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Note that for the an undirected graph,  $A_{ij} = A_{ji}$  so the adjacency matrix is a symmetric matrix. Moreover, we can put value attributes to the edges and define  $A_{ij}$  to be:

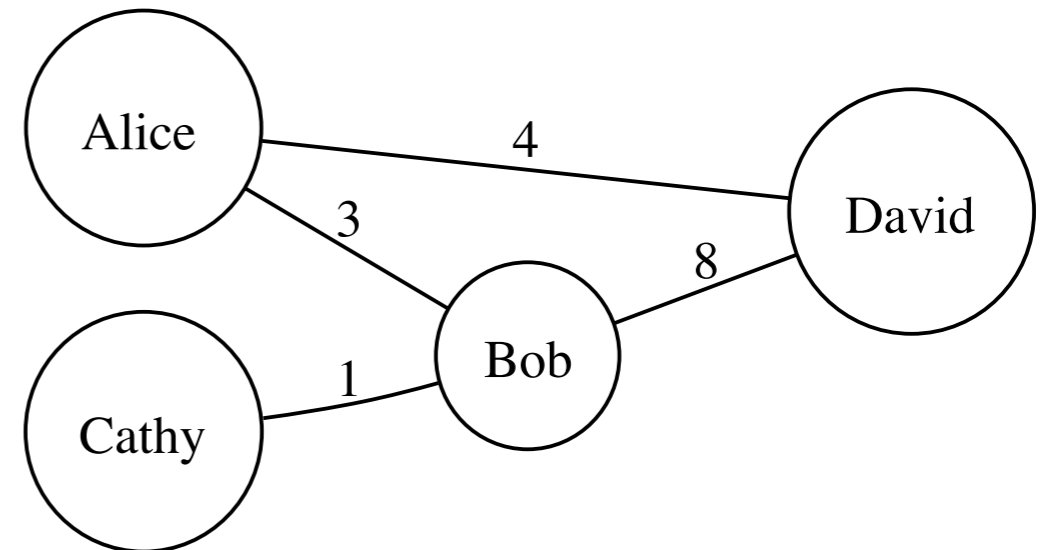
$$A_{ij} = \begin{cases} w, & \text{if there is an edge from } v_i \text{ to } v_j \text{ and } w \text{ is its weight} \\ 0, & \text{if there is no edge from } v_i \text{ to } v_j. \end{cases} \quad (2)$$



# Examples of Adjacency Matrix



	<i>Alice</i>	<i>Bob</i>	<i>Cathy</i>	<i>David</i>
<i>Alice</i>	—	1	0	0
<i>Bob</i>	0	—	0	1
<i>Cathy</i>	0	1	—	0
<i>David</i>	1	0	0	—



	<i>Alice</i>	<i>Bob</i>	<i>Cathy</i>	<i>David</i>
<i>Alice</i>	—	3	0	4
<i>Bob</i>	3	—	1	8
<i>Cathy</i>	0	1	—	0
<i>David</i>	4	8	0	—



# Define Length, Path, and Cycle

- A **path**  $p$  in  $G$  is a sequence of vertices  $w_1, w_2, w_3, \dots, w_N$  such that  $(w_i, w_{i+1}) \in E$  for  $1 \leq i \leq N$  and that  $w_i \neq w_j$  with  $i \neq j$ .
- The length of  $p$  is the number of edges on the path, which is equal to  $N - 1$ .
- The length can be zero for the case of a single vertex.
- The distance between two nodes is the length of shortest path.
- A path with no repeated vertices is called a *simple path*.
- A cycle with no repeated vertices aside from the starting and ending vertex is a simple cycle. A simple cycle that includes every vertex of the graph is known as a Hamiltonian cycle.
- Two paths are independent if they do not have any internal vertex in common.
- For a weighted graph, the weight of a path is the sum of the weights of the traversed edges.



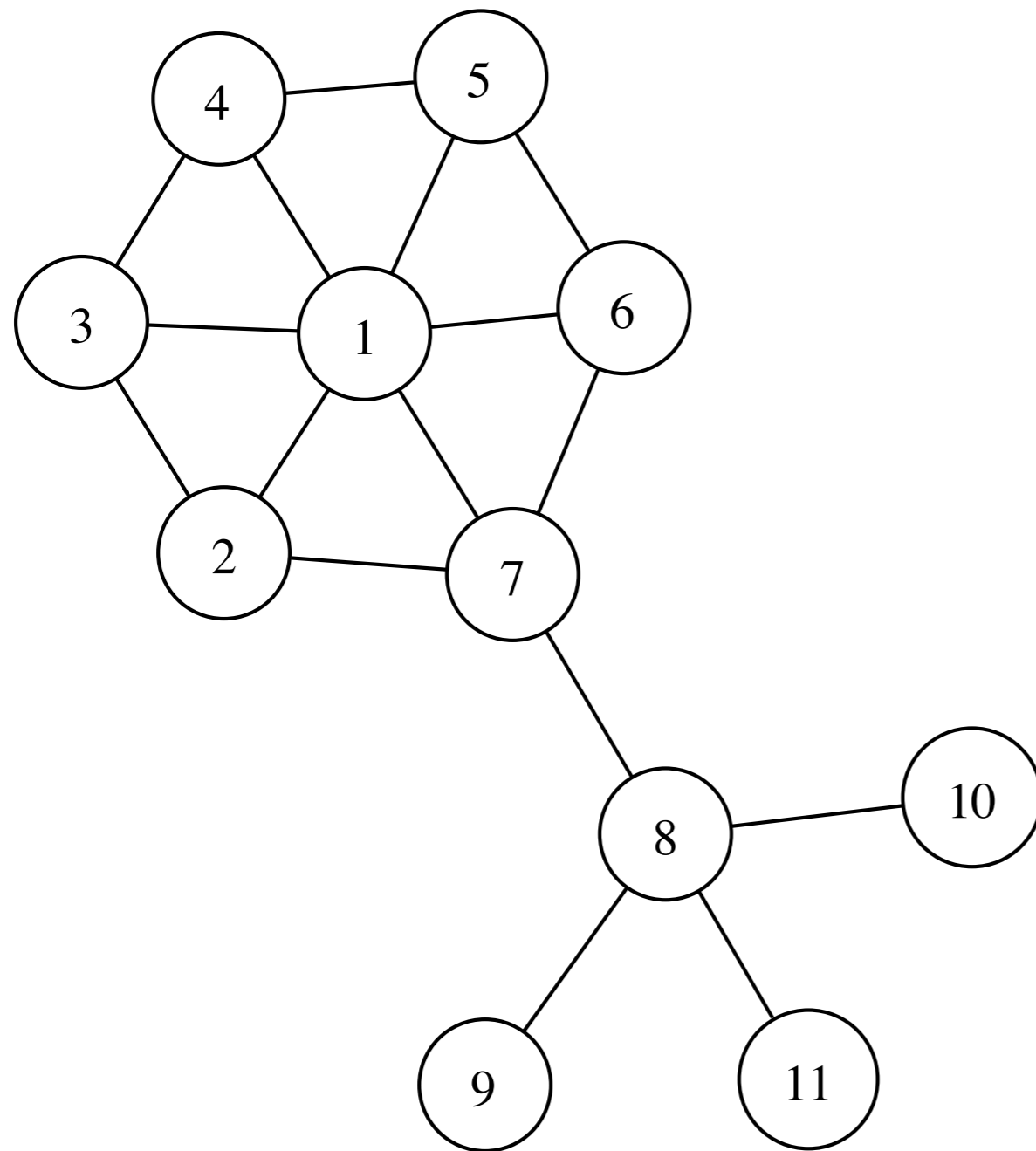


# Define Trail and Walk

- A **trail**  $t$  in  $G$  is a sequence of vertices  $w_1, w_2, w_3, \dots, w_N$  such that  $(w_i, w_{i+1}) \in E$  for  $1 \leq i \leq N$  and that  $e_i \neq e_j$  with  $i \neq j$ .
- A **walk** is an alternating sequence of vertices and edges, beginning and ending with a vertex, where each vertex is incident to both the edge that precedes it and the edge that follows it in the sequence, and where the vertices that precede and follow an edge are the end vertices of that edge.
- A walk is *closed* if its first and last vertices are the same, and *open* if they are different.



# Example of Path, Trail, and Walk



Walk is the most general!

Walk:  $\{1,4,5,1,6,5,4,3\}$

Trail is a special type of walk with no repeated edges.

Trail:  $\{2,7,6,1,7,8,9\}$

Path is a walk with no repeated vertices.

Path:  $\{1,4,5,6,7,8,9\}$

A walk is closed if the starting and ending nodes are the same.

A cycle is a closed trail. A cycle of length  $k$  is called a  $k$ -cycle.



# Define Some Graph Properties

- The **eccentricity**  $\epsilon$  of a vertex  $v$  is the greatest distance between  $v$  and any other vertex.
- The **radius** of a graph is the minimum eccentricity of any vertex.
- The **diameter** of a graph is the maximum eccentricity of any vertex in the graph, i.e., it is the greatest distance between any two vertices.
- A **peripheral vertex** in a graph of diameter  $d$  is one that is distance  $d$  from some other vertex, i.e., a vertex that achieves the diameter.
- A **pseudo-peripheral vertex**  $v$  has the property that for any vertex  $u$ , if  $v$  is as far away from  $u$  as possible, then  $u$  is as far away from  $v$  as possible.
- The **girth** of a graph is the length of a shortest cycle contained in the graph. If the graph does not contain any cycles, its girth is defined to be infinity.



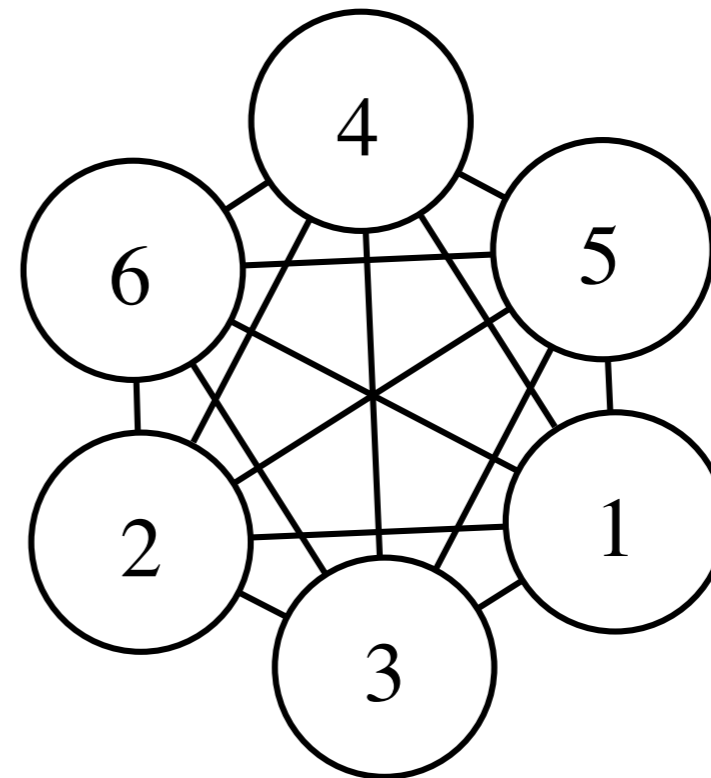
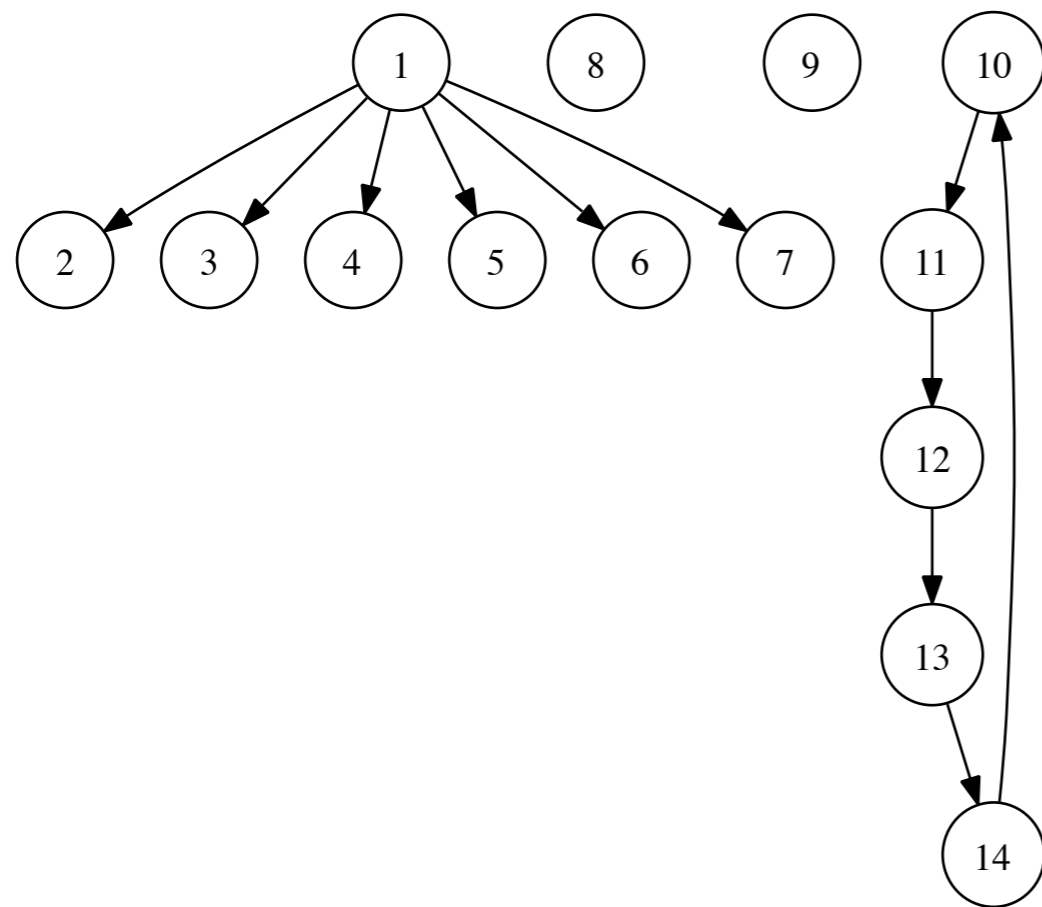


# Define Connected Component

- A **connected component** of an undirected graph is a subgraph in which any two vertices are connected to each other by paths, and to which no more vertices or edges can be added while preserving its connectivity. That is, it is a maximal connected subgraph.
- An undirected graph is connected if there is a path from every vertex to every other vertex.
- A directed graph with this property is called **strongly connected**.
- A **weakly connected graph** is a directed graph which is not strongly connected, but the underlying graph (without direction to the edges) is connected.
- A **complete graph** is a graph in which there is an edge between every pair of vertices.

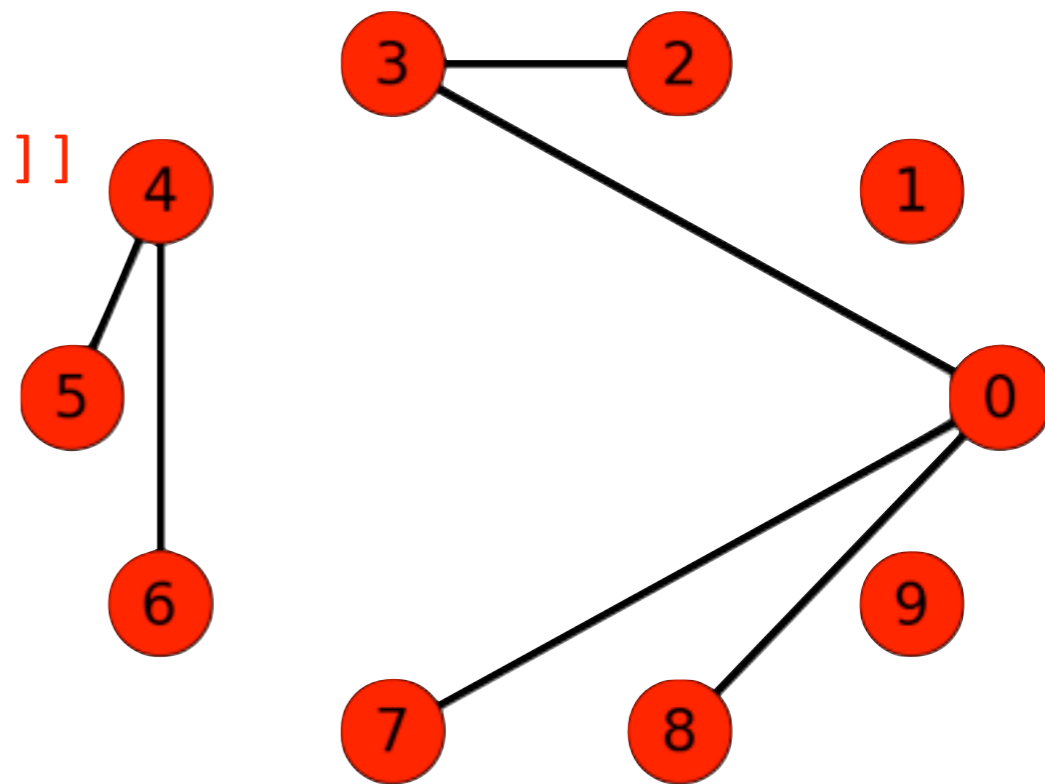


# Example of Components



# Connected Components

```
>>> G =  
nx.generators.random_graphs.gnp_random_graph(10,0.15)  
>>>  
>>> nx.is_connected(G)  
False  
>>> nx.number_connected_components(G)  
4  
>>> nx.connected_components(G)  
[[0, 8, 2, 3, 7], [4, 5, 6], [1], [9]]
```





# Define Cutpoint

- A **cutpoint** is a vertex whose removal from the graph increases the number of components. That is, it makes some points unreachable from some others. It disconnects the graph.
- A **cutset** is a collection of points whose removal increases the number of components in a graph.
- A **minimum weight cutset** consists of the smallest set of points that must be removed to disconnect a graph. The number of points in a minimum weight cutset is called the **point connectivity** of a graph.
- If a graph has a cutpoint, the connectivity of the graph is 1.
- The minimum number of points separating two nonadjacent points  $s$  and  $t$  is also the maximum number of point-disjoint paths between  $s$  and  $t$ .

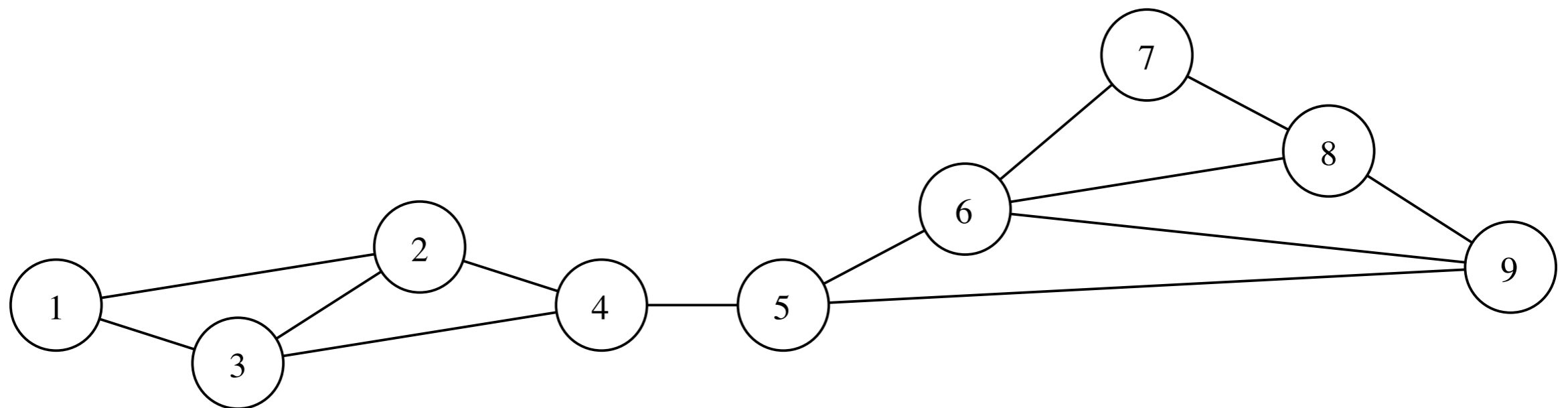
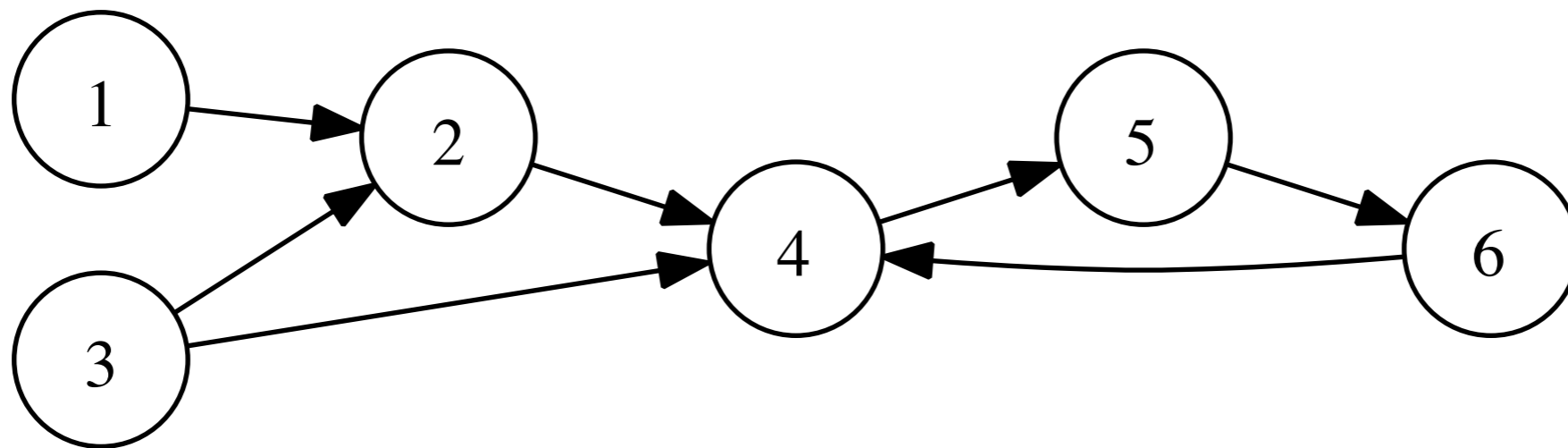


# Define Cutpoint

- A **bridge** is an edge whose removal from a graph increases the number of components (disconnects the graph).
- An **edge cutset** is a collection of edges whose removal disconnects a graph.
- A local bridge of degree  $k$  is an edge whose removal causes the distance between the endpoints of the edge to be at least  $k$ .
- The **edge-connectivity** of a graph is the minimum number of lines whose removal would disconnect the graph. The minimum number of edges separating two nonadjacent points  $s$  and  $t$  is also the maximum number of edge-disjoint paths between  $s$  and  $t$ .



# Example of a Cutpoint and Bridge



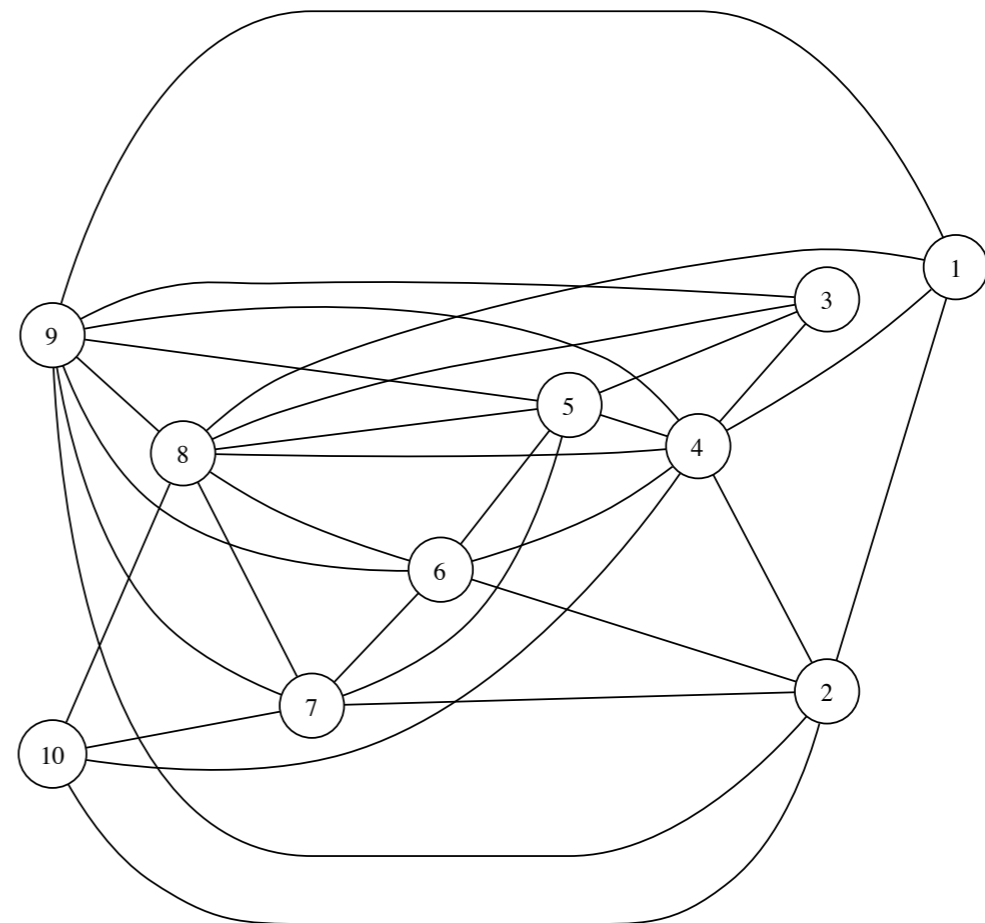
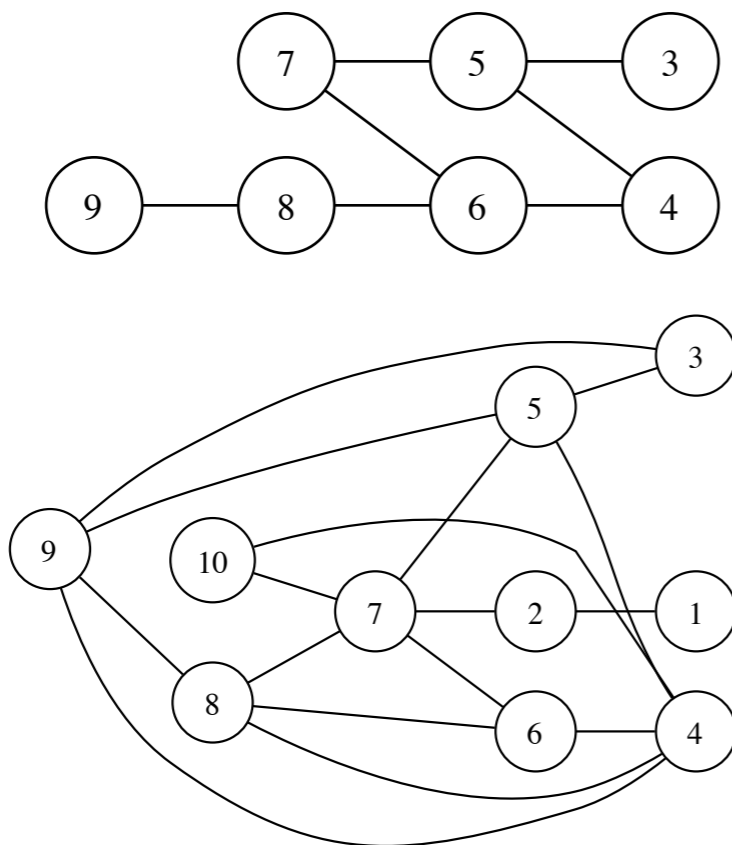


# Define Graph Density

For undirected simple graphs, the **graph density** is defined as:

$$D = \frac{2|E|}{|V|(|V| - 1)}, \quad (1)$$

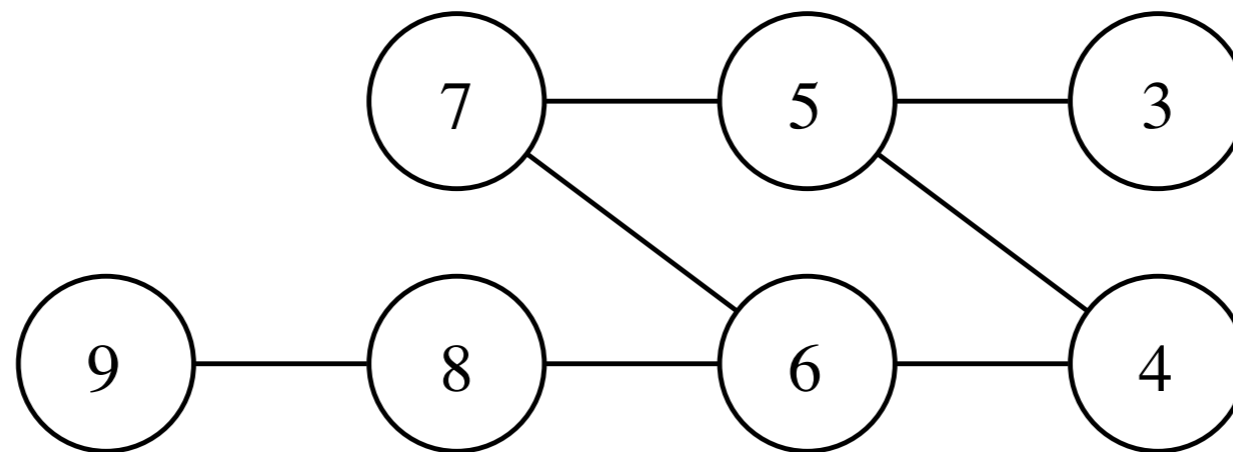
where  $|E|$  denotes the number of edges and  $|V|$  denotes the number of vertices. The maximum number of edges is  $\frac{1}{2}|V|(|V| - 1)$ , so the maximal density is 1 (for complete graphs) and the minimal density is 0.



# Define Graph Distance

- The **distance**  $d_G(u, v)$  between two (not necessary distinct) vertices  $u$  and  $v$  in a graph  $G$  is the length of a shortest path between  $u$  and  $v$ .
- When  $u$  and  $v$  are identical, their distance is 0. When  $u$  and  $v$  are unreachable from each other, their distance is defined to be infinity  $\infty$ .
- The average distance is the summation of the distance between all pairs of reachable nodes divided by the number of nodes.

$$d_{av}(G) = \frac{\sum_{u,v}^V d_G(u, v)}{|V|}$$



# Define Degree Centrality

Let  $G = (V, E)$  with  $n$  vertices, the **Degree Centrality**  $C_D(v)$  for a vertex  $v$  is defined as

$$C_D(v) = \frac{\deg(v)}{n - 1} \quad (1)$$

For directed graphs, the above can be decomposed to include indegree and outdegree as

$$C_{Din}(v) = \frac{\text{indeg}(v)}{n - 1} \quad (2)$$

$$C_{Dout}(v) = \frac{\text{outdeg}(v)}{n - 1} \quad (3)$$





# Define Group Degree Centralization

The **Group Degree Centralization** is defined by Freeman as

$$C_D(G) = \frac{\sum_v (\Delta_G - C_D(v))}{\max_H \sum_{v \in H} (\Delta_H - C_D(v))}, \quad (1)$$

where  $\Delta_G$  is the maximum degree of any node in  $G$ ,  $C_D(v)$  is the degree of node  $v$  in  $G$  and the maximum is taken over all possible graph of the same order (the same number of nodes).



# Define Group Degree Centralization

Let  $n$  and  $m$  denote the numbers of nodes and edges, respectively. We have

$$C_D(G) = \frac{n\Delta_G - \sum_v (C_D(v))}{(n-1)(n-2)}. \quad (1)$$

For an undirected graph,

$$C_D(G) = \frac{n\Delta_G - 2m}{(n-1)(n-2)} \quad (2)$$

For a directed graph,

$$C_D^{in}(G) = \frac{n\Delta_G^{in} - m}{(n-1)(n-2)} \quad (3)$$

and

$$C_D^{out}(G) = \frac{n\Delta_G^{out} - m}{(n-1)(n-2)}. \quad (4)$$



# Define Degree Centrality for $G$

Let  $V^*$  be the node with the highest degree centrality in  $G$ . Let  $G' = (V', E')$  be the  $n$  node connected graph that maximizes the following quantity

$$H = \sum_{j=1}^{|V'|} C_D(v'^*) - C_D(v'_j) \quad (1)$$

Then the degree centrality of the graph  $G$  is defined as

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v^*) - C_D(v_i)]}{H} \quad (2)$$

$H$  is maximized when the graph  $G'$  contains one node that is connected to all other nodes are connected only to this one central node (a star graph). In this case

$$H = (n - 1) \left(1 - \frac{1}{n - 1}\right) = n - 1 \quad (3)$$

so the degree centrality of  $G$  reduces to

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v^*) - C_D(v_i)]}{n - 1} \quad (4)$$



# Define Closeness Centrality

Let  $d(u, v)$  denote the distance from  $u$  to  $v$  and  $D(v) = \sum_u d(v, u)$  be the total distance from  $v$  to all other nodes. The **Closeness Centrality** of  $v$  is measured by  $1/D(v)$  and normalized to  $C_C(v) = (n - 1)/D(v)$  since the minimum  $D(v)$  is  $n - 1$ , which happens at the center of a star graph. Freeman defines the group centrality as follows,

$$C_C(G) = \frac{\sum_v (C_C(v^*) - C_C(v))}{\max_H \sum_{v \in H} (C_C(v^*) - C_C(v))}, \quad (1)$$

where  $v^*$  is the node of maximum closeness.





# Define Between Centrality

The **Between Centrality** is a measure of a vertex within a graph (this can also be extended to edge as well). Vertices that occur on many shortest paths between other vertices have higher betweenness than those that do not. Hence, Betweenness Centrality of a node counts the number of times that a node lies along the shortest path between two others vertices in the graph. It is defined as

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{g\sigma_{st}}. \quad (1)$$

where  $\sigma_{st}$  is the number of shortest paths from  $s$  to  $t$  and  $\sigma_{st}(v)$  is the number of shortest paths from  $s$  to  $t$  that pass through a vertex  $v$ .

The normalized betweenness of undirected graphs is given by

$$C'_B(v) = \frac{C_B(v)}{(n-1)(n-2)/2}. \quad (2)$$



# Define Group Between Centrality

The normalized betweenness of directed graphs is given by

$$C'_B(v) = \frac{C_B(v)}{(n-1)(n-2)}. \quad (1)$$

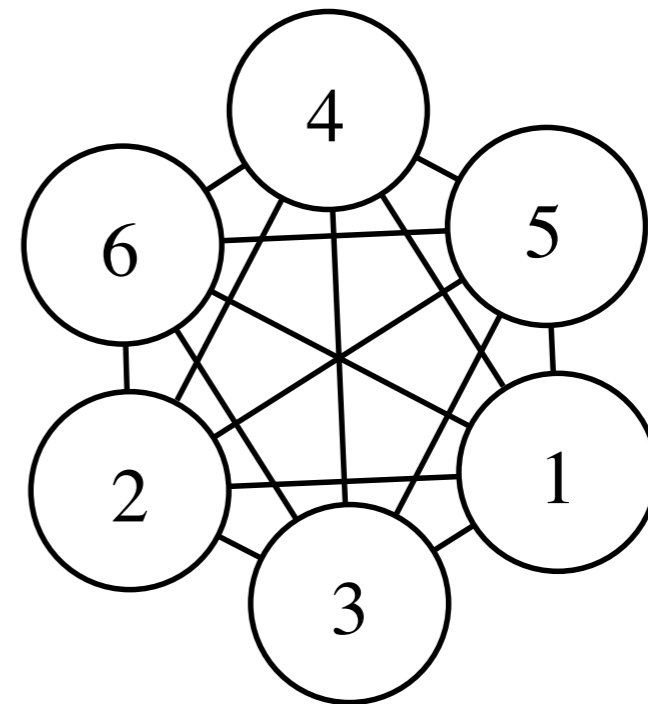
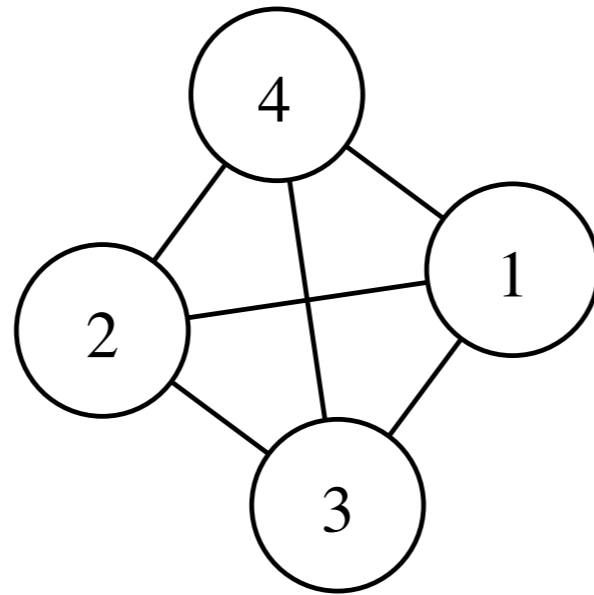
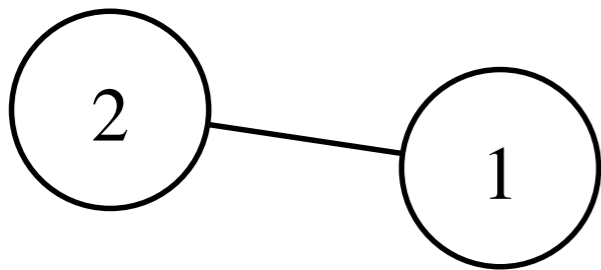
To compute the group betweenness centrality, we compute the one of a star at first. For a star, the center has betweenness  $(n-1)(n-2)/2$  and it is zero for all the others. The group centrality of a star is then  $(n-1)^2(n-2)/2$ . Then we have

$$\begin{aligned} C_B(G) &= \frac{\sum_v (C_B(v^*) - C_B(v))}{(n-1)^2(n-2)/2} \\ &= \frac{\sum_v (C'_B(v^*) - C'_B(v))}{n-1} \end{aligned} \quad (2)$$



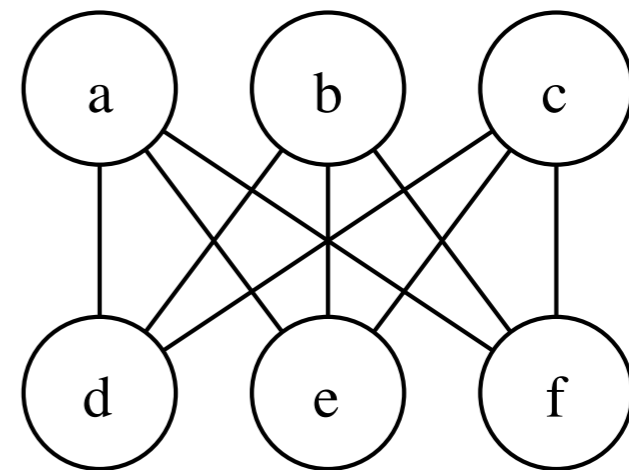
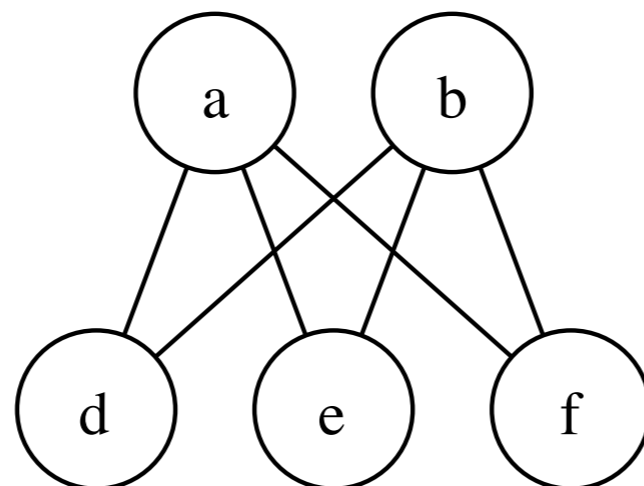
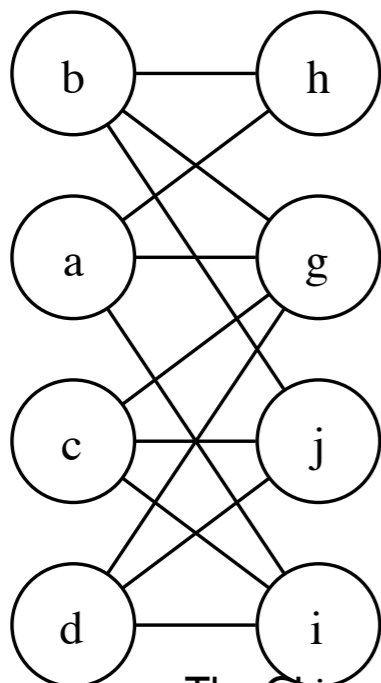
# Complete Graph

- A **complete graph** is a simple graph in which every pair of distinct vertices is connected by an edge.
- The complete graph on  $n$  vertices has  $n$  vertices and  $n(n - 1)/2$  edges, and is denoted by  $K_n$ .
- It is a regular graph of degree  $n - 1$ .
- All complete graphs are their own cliques. They are maximally connected as the only vertex cut which disconnects the graph is the complete set of vertices.



# Bipartite Graph

- A **bipartite graph** (or bigraph) is a graph whose vertices can be divided into two disjoint sets  $U$  and  $V$  such that every edge connects a vertex in  $U$  to one in  $V$ , i.e.,  $U$  and  $V$  are independent sets.
- Equivalently, a bipartite graph is a graph that does not contain any odd-length cycles.
- A **balanced bipartite graph** is a bipartite graph that satisfy the condition  $|U| = |V|$ .
- A **complete bipartite graph**  $G = (U + V, E)$  is bipartite such that for any two vertices  $u \in U$  and  $v \in V$  that  $(u, v)$  is an edge in  $G$ .
- The complete bipartite graph with partitions of size  $|U| = m$  and  $|V| = n$ , is denoted  $K_{m,n}$ .





# Properties of Bipartite Graphs

- A graph is bipartite if and only if it does not contain an odd cycle. Therefore, a bipartite graph cannot contain a clique of size 3 or more.
- A graph is bipartite if and only if it is 2-colorable.
- The size of minimum vertex cover is equal to the size of the maximum matching.
- The size of the maximum independent set plus the size of the maximum matching is equal to the number of vertices.



# Basic Functions

`is_bipartite(G)`

- Returns True if graph  $G$  is bipartite, False if not.
- `is_bipartite_node_set(G, nodes)` Returns True if nodes and  $G/nodes$  are a bipartition of  $G$ .

`sets(G)`

- Returns bipartite node sets of graph  $G$ .

`color(G)`

- Returns a two-coloring of the graph.

`density(B, nodes)`

- Return density of bipartite graph  $B$ . `degrees(B, nodes[, weighted])`  
Return the degrees of the two node sets in the bipartite graph  $B$ .

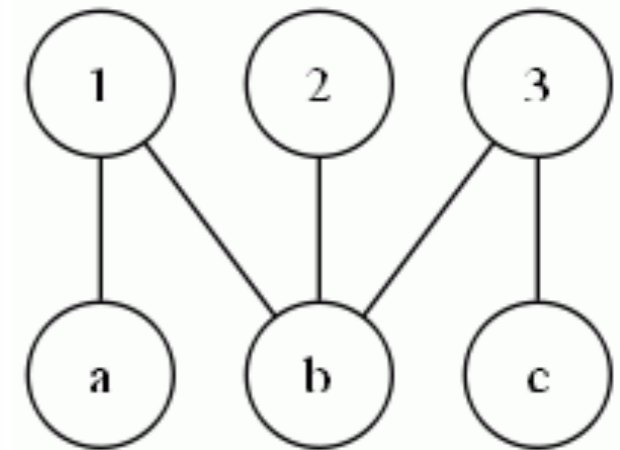


# Bipartite Module in NetworkX

- This module provides functions and operations for bipartite graphs. Bipartite graphs  $G(X, Y, E)$  have two node sets  $X, Y$  and edges in  $E$  that only connect nodes from opposite sets.

- For example:

```
>>> import networkx as nx
>>> top_nodes=[1,1,2,3,3]
>>> bottom_nodes=['a','b','b','b','c']
>>> edges=zip(top_nodes,bottom_nodes) # create 2-tuples of
edges
>>> B=nx.Graph(edges)
>>> print(B.edges())
```



- The bipartite algorithms are not imported into the networkx (version 1.5) namespace at the top level so you need to do:

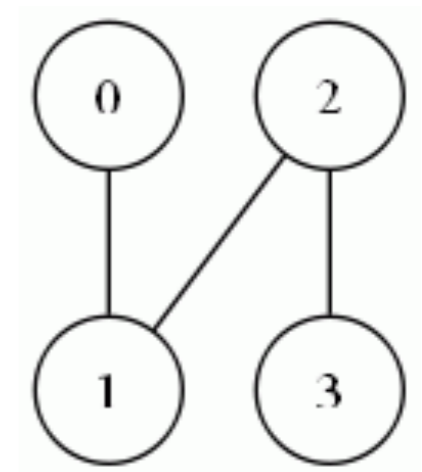
```
>>> from networkx.algorithms import bipartite
```



# Examples of Basic Functions

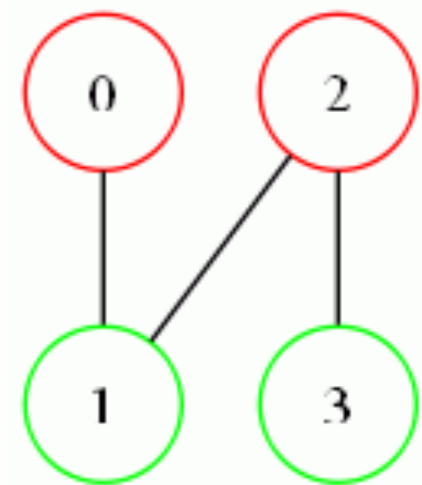
- `networkx.algorithms.bipartite.basic.sets`

```
>>> from networkx.algorithms import bipartite
>>> G = nx.path_graph(4)
>>> X, Y = bipartite.sets(G)
>>> list(X)
[0, 2]
>>> list(Y)
[1, 3]
```



- `networkx.algorithms.bipartite.basic.color`

```
>>> from networkx.algorithms import bipartite
>>> G = nx.path_graph(4)
>>> c = bipartite.color(G)
>>> print(c)
{0: 1, 1: 0, 2: 1, 3: 0}
```

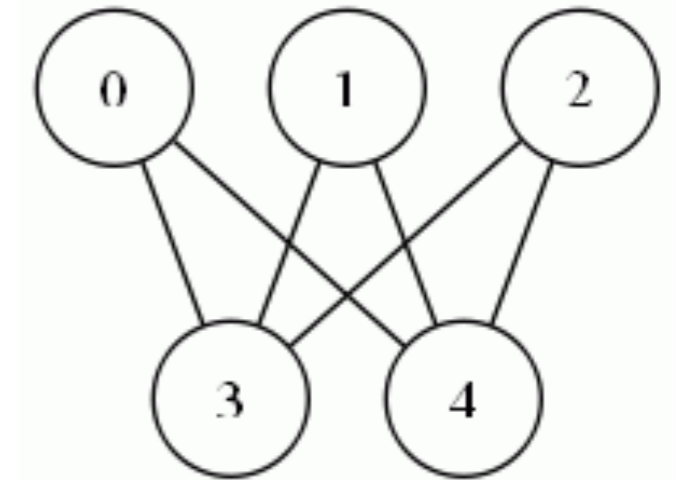




# More Examples

- `networkx.algorithms.bipartite.basic.density`

```
>>> from networkx.algorithms import bipartite
>>> G = nx.complete_bipartite_graph(3,2)
>>> X=set([0,1,2])
>>> bipartite.density(G,X)
1.0
>>> Y=set([3,4])
>>> bipartite.density(G,Y)
1.0
```



- `networkx.algorithms.bipartite.basic.degrees`

```
>>> from networkx.algorithms import bipartite
>>> G = nx.complete_bipartite_graph(3,2)
>>> Y=set([3,4])
>>> degX,degY=bipartite.degrees(G,Y)
>>> degX
{0: 2, 1: 2, 2: 2}
```



# Other Functions

- **Spectral**

`spectral_bipartivity(G[, nodes, weight])`

- Returns the spectral bipartivity.

- **Clustering**

`clustering(G[, nodes, mode])`

- Compute a bipartite clustering coefficient for nodes.

`average_clustering(G[, nodes, mode])`

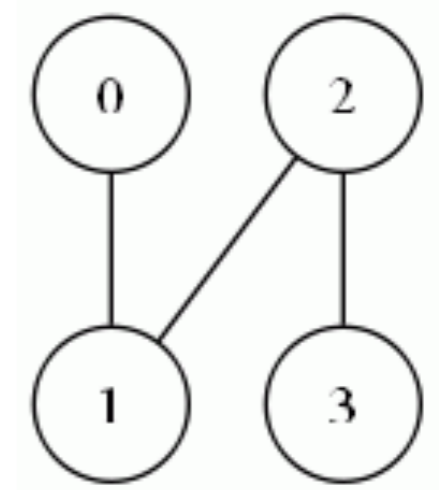
- Compute the average bipartite clustering coefficient.



# Examples of Clustering

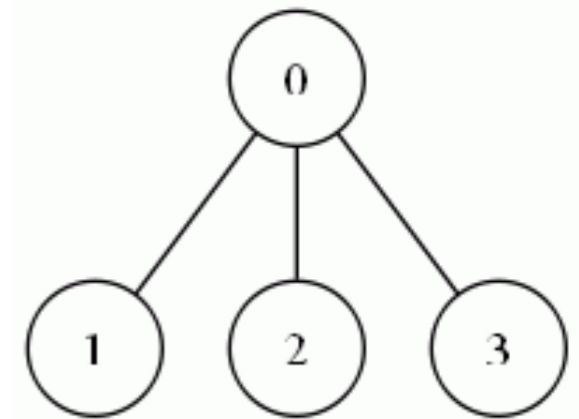
- `networkx.algorithms.bipartite.cluster.clustering`

```
>>> from networkx.algorithms import bipartite
>>> G=nx.path_graph(4) # path is bipartite
>>> c=bipartite.clustering(G)
>>> c[0]
0.5
>>> c=bipartite.clustering(G,mode='min')
>>> c[0]
1.0
```



- `networkx.algorithms.bipartite.cluster.average_clustering`

```
>>> from networkx.algorithms import bipartite
>>> G=nx.star_graph(3) # path is bipartite
>>> bipartite.average_clustering(G)
0.75
>>> X,Y=bipartite.sets(G)
>>> bipartite.average_clustering(G,X)
0.0
>>> bipartite.average_clustering(G,Y)
1.0
```



# Bipartite Cluster Clustering

- The bipartite clustering coefficient is a measure of local density of connections defined as

$$c_u = \frac{\sum_{v \in N(N(u))} c_{uv}}{N(N(u))}$$

where  $N(N(u))$  are the second order neighbors of  $u$  in  $G$  excluding  $u$ , and  $c_{uv}$  is the pairwise clustering coefficient between nodes  $u$  and  $v$ .

- $c_{uv}$  can be defined in three ways.

–

$$c_{uv} = \frac{\|N(u) \cap N(v)\|}{\|N(u) \cup N(v)\|}$$

– min:

$$c_{uv} = \frac{\|N(u) \cap N(v)\|}{\min(\|N(u) \cup N(v)\|)}$$

– max:

$$c_{uv} = \frac{\|N(u) \cap N(v)\|}{\max(\|N(u) \cup N(v)\|)}$$





# More Functions

- **Redundancy**

`node_redundancy(G[, nodes])`

- Compute bipartite node redundancy coefficient.

- **Centrality**

`closeness centrality(G, nodes[, normalized])`

- Compute the closeness centrality for nodes in a bipartite network.

`degree centrality(G, nodes)`

- Compute the degree centrality for nodes in a bipartite network.

`betweenness centrality(G, nodes)`

- Compute betweenness centrality for nodes in a bipartite network.



# Examples of Redundancy

- `networkx.algorithms.bipartite.redundancy.node_redundancy`

```
>>> from networkx.algorithms import bipartite
>>> G = nx.cycle_graph(4)
>>> rc = bipartite.node_redundancy(G)
>>> rc[0]
1.0
```



# Example

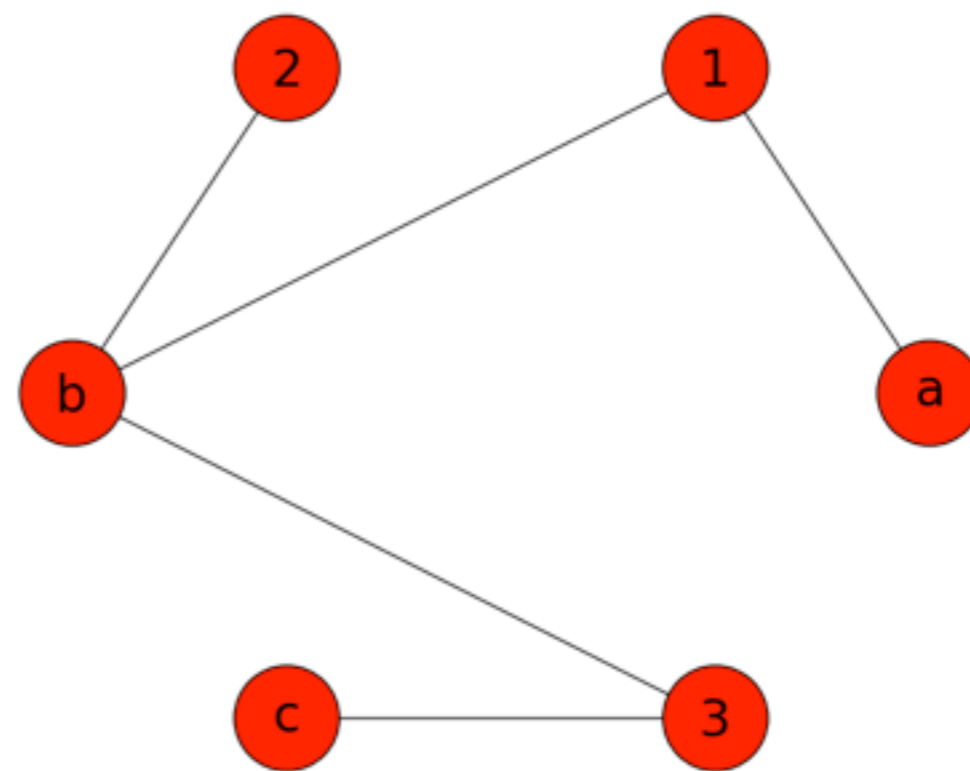
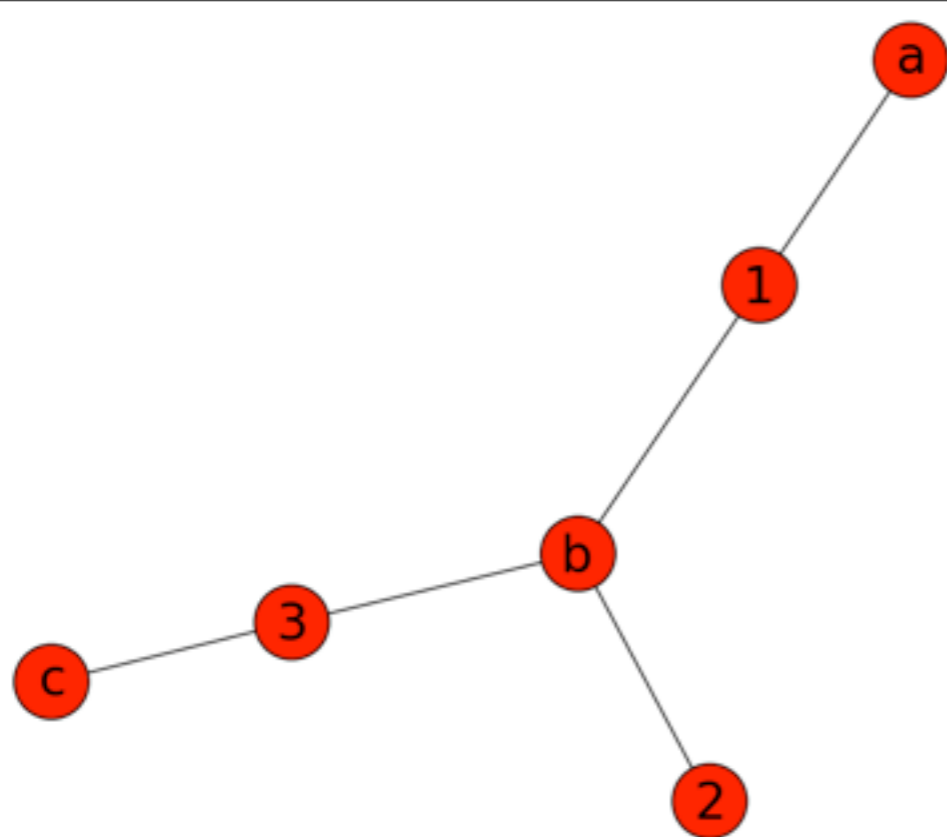
```
import networkx as nx
import matplotlib.pyplot as plt
import pygraphviz

top_nodes=[1,1,2,3,3]
bottom_nodes=['a','b','b','b','c']
edges=zip(top_nodes,bottom_nodes) # create 2-tuples of
edges

G=nx.Graph(edges)
print(G.edges())

nx.draw(G)
plt.savefig("example.png")
plt.show()
```



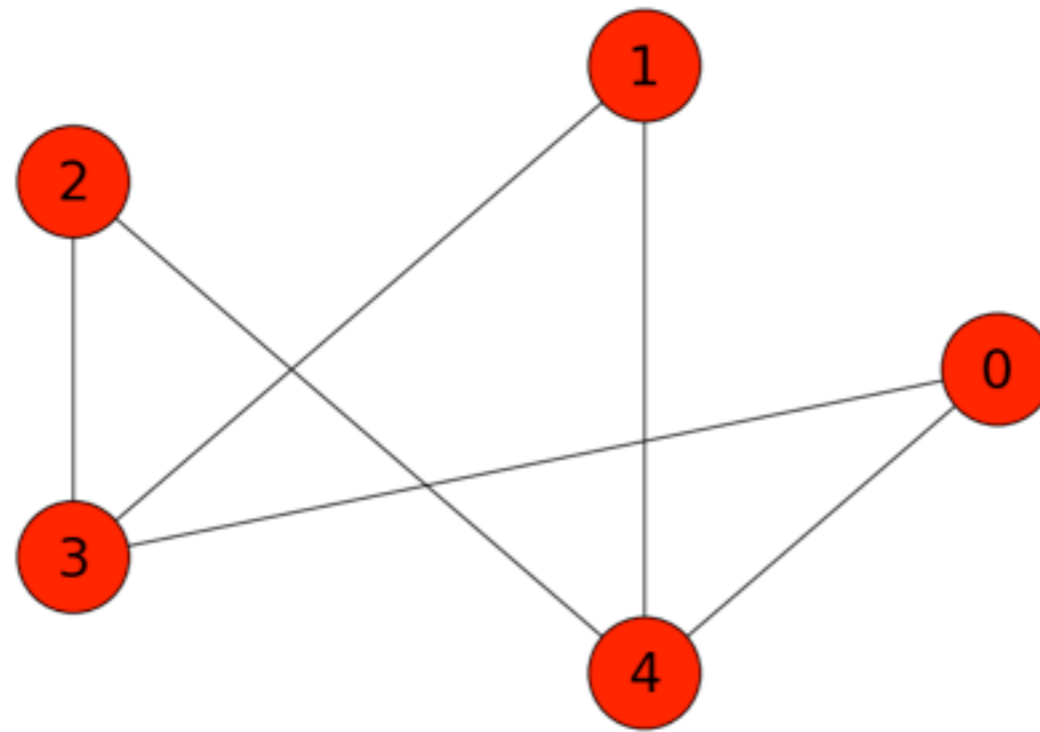


```

>>> centrality.degree Centrality(G)
{'a': 0.20000000000000000001, 1: 0.40000000000000000002, 2:
0.20000000000000000001, 'b': 0.60000000000000000009, 'c':
0.20000000000000000001, 3: 0.40000000000000000002}
>>> centrality.betweenness Centrality(G)
{'a': 0.0, 1: 0.40000000000000000002, 2: 0.0, 'b':
0.80000000000000000004, 'c': 0.0, 3: 0.40000000000000000002}
>>> centrality.closeness Centrality(G)
{'a': 0.38461538461538464, 1: 0.55555555555555555558, 2:
0.4545454545454545453, 'b': 0.7142857142857143, 'c':
0.38461538461538464, 3: 0.55555555555555555558}
  
```







```
>>> from networkx.algorithms import bipartite
>>> G = nx.complete_bipartite_graph(3,2)
>>> X=set([0,1,2])
>>> bipartite.density(G,X)
>>> Y=set([3,4])
```

```
>>> centrality.degree centrality(G)
{0: 0.5, 1: 0.5, 2: 0.5, 3: 0.75, 4: 0.75}
>>> centrality.betweenness centrality(G)
{0: 0.055555555555555552, 1: 0.055555555555555552, 2:
0.055555555555555552, 3: 0.25, 4: 0.25}
>>> centrality.closeness centrality(G)
{0: 0.66666666666666663, 1: 0.66666666666666663, 2:
0.66666666666666663, 3: 0.80000000000000004, 4: 0.80000000000000004}
```



# Subgroup Cohesion

- A **Clique** in an undirected graph  $G = (V, E)$  is a subset of the vertex set  $C \subseteq V$ , such that for every two vertices in  $C$ , there exists an edge connecting the two. This is equivalent to saying that the subgraph induced by  $C$  is complete.
- The size of the clique is the number of vertices it contains.
- The **clique number**  $\omega(G)$  of a graph  $G$  is the order of a largest clique in  $G$ .
- An  **$n$ -clique**  $S$  of a graph is a maximal set of nodes in which for all  $u, v \in S$ , the graph-theoretic distance  $d(u, v) \leq n$ .
- In other words, an  $n$ -clique is a set of nodes in which every node can reach every other in  $n$  or fewer steps, and the set is maximal in the sense that no other node in the graph is distance  $n$  or less from every other node in the subgraph.
- A 1-clique is the same as an ordinary clique.



# Clan

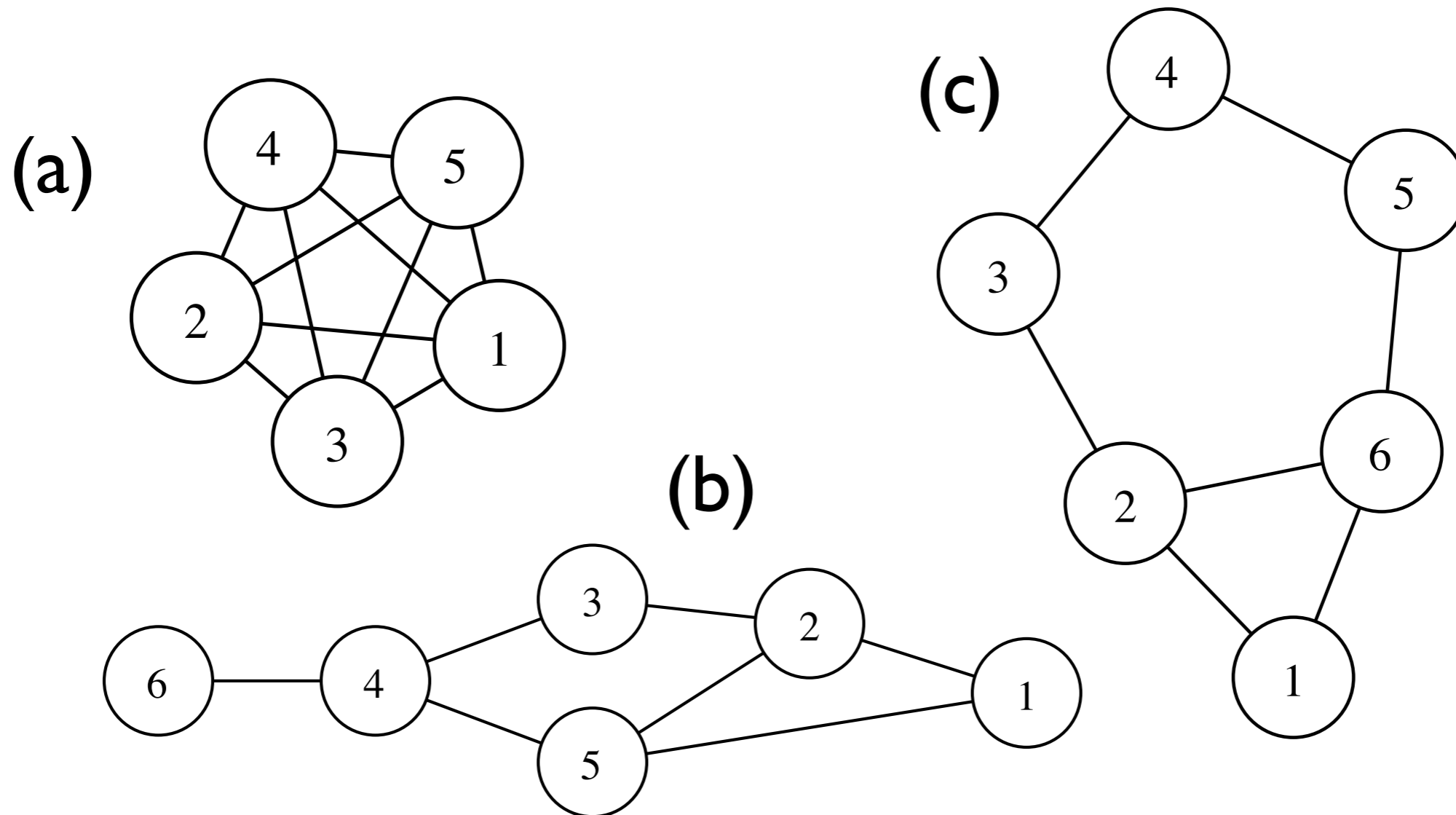
An  **$n$ -clan** is an  $n$ -clique in which the diameter of the subgraph  $G'$  induced by  $S$  is less than or equal to  $n$ .

An  **$n$ -club** is a subset  $S$  of nodes such that in the subgraph induced by  $S$ , the diameter is  $n$  or less. Every  $n$ -clan is both an  $n$ -club and an  $n$ -clique.

A  **$k$ -plex** is a subset  $S$  of nodes such that every member of the set is connected to  $n - k$  others, where  $n$  is the size of  $S$ . The  **$k$ -plex** generalizes the clique by relaxing density.



# Example of Cliques, Clans, Clubs, etc.



(a) A complete graph and also a clique of size 5. (b) An example of a clique of size 3. (c) An example of 2-clique with  $\{1, 2, 3, 4, 5\}$ . An example of 2-clan with  $\{2, 3, 4, 5, 6\}$ . An example of 2-club with  $\{1, 2, 3, 6\}$ .





# The Clique Problem

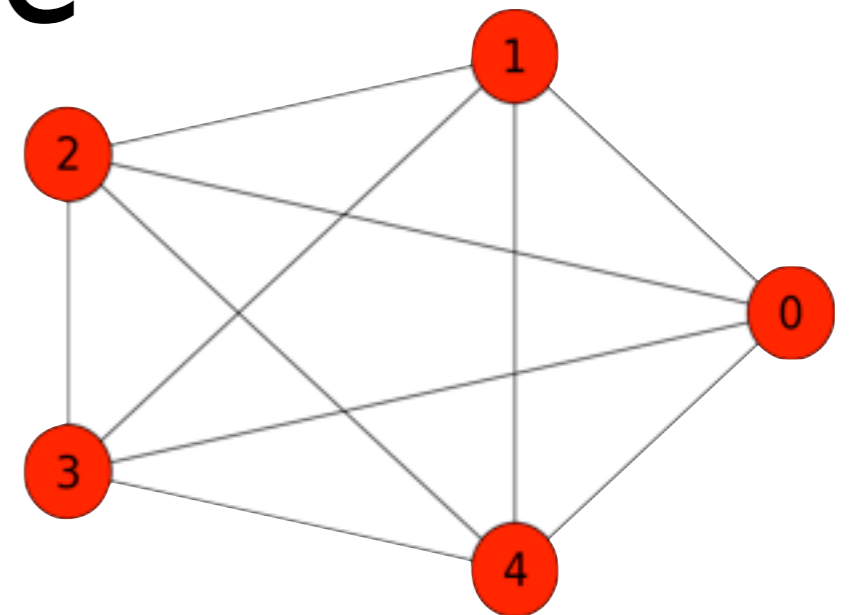
- There is a clique of size at least  $k$  iff there is an independent set of size at least  $k$  in the complement graph.
- Brute Force Algorithm
  - Examine each subgraph with at least  $k$  vertices and check to see if it forms a clique.
  - Polynomial if  $k$  is the number of vertices, or a constant

$$\binom{V}{k} = \frac{V!}{k!(V-k)!}$$

- Consider each node to be a clique of size one, and to merge cliques into larger cliques
- Linear time by the edges
- Disjoint-set data structure



# Clique Example



```
from networkx.algorithms import clique
```

```
G = nx.complete_graph(5)
```

```
clique.graph_clique_number(G) # Return the clique number  
(size of the largest clique) for G
```

```
5
```

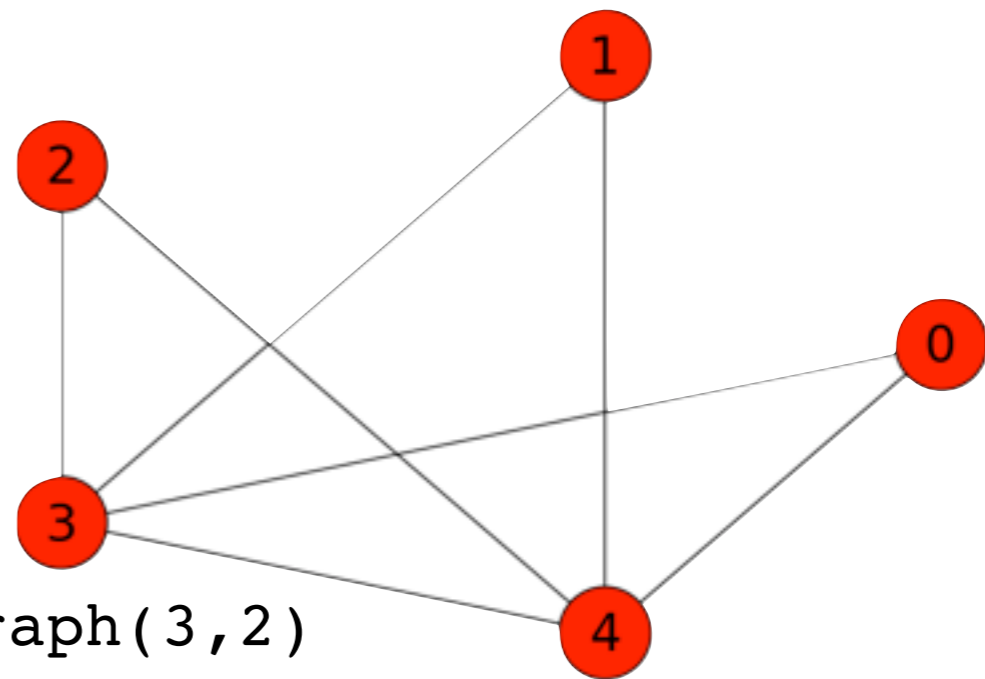
```
list(clique.find_cliques(G)) # Search for all maximal  
cliques in a graph.
```

```
[[0, 1, 2, 3, 4]]
```

```
clique.cliques_containing_node(G) # Returns a list of  
cliques containing the given node.
```

```
{0: [[0, 1, 2, 3, 4]], 1: [[0, 1, 2, 3, 4]], 2: [[0, 1, 2,  
3, 4]], 3: [[0, 1, 2, 3, 4]], 4: [[0, 1, 2, 3, 4]]}
```





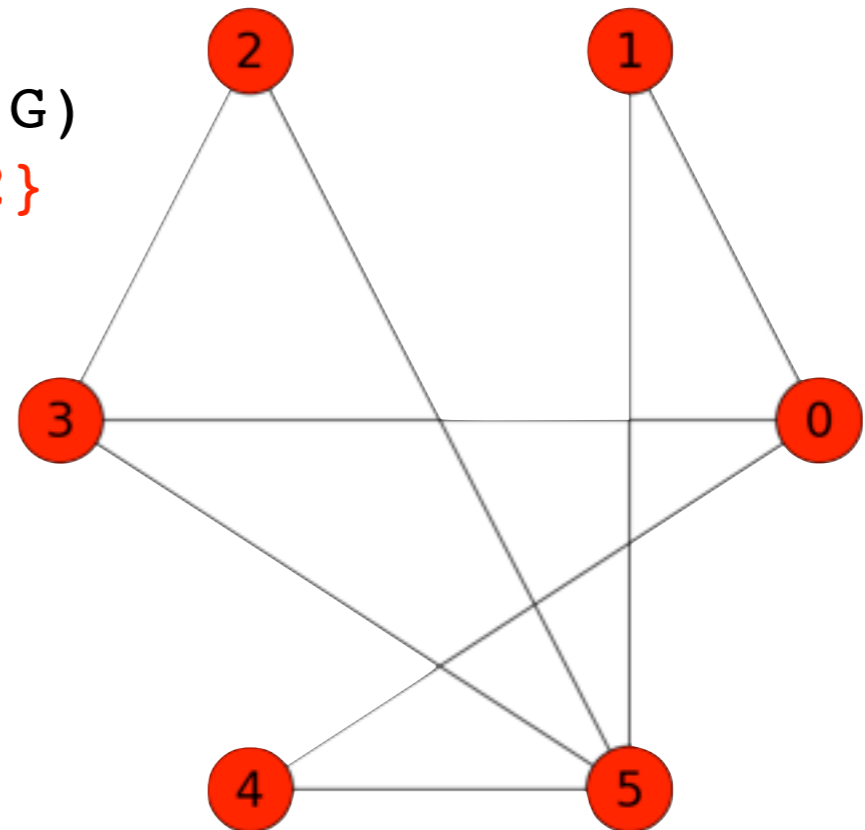
```
>>> G = nx.complete_bipartite_graph(3,2)
>>> G.add_edge(3,4)
>>>
>>> clique.graph_clique_number(G) # Return the clique
number (size of the largest clique) for G
3
>>> list(clique.find_cliques(G)) # Search for all maximal
cliques in a graph.
[[3, 4, 0], [3, 4, 1], [3, 4, 2]]
>>> clique.cliques_containing_node(G) # Returns a list of
cliques containing the given node.
{0: [[3, 4, 0]], 1: [[3, 4, 1]], 2: [[3, 4, 2]], 3: [[3, 4,
0], [3, 4, 1], [3, 4, 2]], 4: [[3, 4, 0], [3, 4, 1], [3, 4,
2]]}
```



# Graph Measures

```
>>> from networkx.algorithms import generators
>>> from networkx.algorithms import distance_measures
>>> G = nx.generators.random_graphs.gnp_random_graph(6,0.5)

>>> distance_measures.diameter(G)
2
>>> distance_measures.eccentricity(G)
{0: 2, 1: 2, 2: 2, 3: 2, 4: 2, 5: 2}
>>> distance_measures.center(G)
[0, 1, 2, 3, 4, 5]
>>> distance_measures.periphery(G)
[0, 1, 2, 3, 4, 5]
>>> distance_measures.radius(G)
2
```





# References

- NetworkX, <http://networkx.lanl.gov/>
- D. J. Cook and L. B. Holder, *Mining Graph Data*, 1st ed. Wiley-Interscience, 2006
- T. G. Lewis, *Network Science: Theory and Applications*, 1st ed. Wiley, 2009.
- M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*. Back Bay Books, 2002.

