


# **Analysis of Social and Information Networks**

Chapter 1 and 2 – Overview and Graphs

# Reddit.com: the frontpage of the Internet













MY SUBREDDITS ▼ POPULAR - ALL - RANDOM - USERS | ASKREDDIT - WORLDNEWS - VIDEOS - FUNNY - TODAYILEARNED - PICS - GAMING - MOVIES - NEWS - GIFS - MILDLYINTERESTING - AWW - SHOWERTHOU

 **reddit** POPULAR **hot** new rising controversial top gilded

Welcome to Reddit,  
the front page of the internet.

[BECOME A REDDITOR](#) and subscribe to one of thousands of communities.

popular in: [Everywhere](#) ▼

-  36.3k [11 different brands of AA batteries, tested in identical flashlights. \[OC\]](#)  (i.redd.it)  
submitted 2 hours ago by [thecrispiestbacon](#) to [r/dataisbeautiful](#)  
2339 comments share save hide report
-  26.7k  [Guinea Pigs...roll out.](#)  (gfycat.com)  
submitted 5 hours ago by [Panda\\_911](#) to [r/Eyebleach](#)  
406 comments share save hide report
-  59.9k  ["You know what you're doing i'll give you that. Excellent form. But how's your footwork?"](#) (i.imgur.com)  
submitted 5 hours ago by [-EG-](#) to [r/gifs](#)  
755 comments share save hide report
-  16.2k  [Inside the blender](#)  (gfycat.com)  
submitted 5 hours ago by [GallowBoob](#) to [r/woahdude](#)  
553 comments share save hide report
-  14.0k  [Mark Hamill just tweeted this](#) (i.redd.it)  
submitted 5 hours ago by [spicy\\_m4ym4ys](#) to [r/wholesomememes](#)  
195 comments share save hide report

# Reddit.com: the frontpage of the Internet

- **Reddit is a website where users can share links to images, blogs, videos, GIFs, etc. and comment on these links**
- **Every signed user can vote on content:**
  - The more upvotes a link/comment has, the higher up it is displayed
  - Downvoted content is pushed further down on the webpage
- **A user-moderated hub for news, hobbies, communities, questions,...**


<https://www.youtube.com/watch?v=tII022aUWQQ>

# Subreddits

- **Subreddits are a smaller community contained in the reddit website**
- **Usually oriented towards one specific topic of interest**
  - r/asoiaf
  - r/TodayILearned
  - r/gaming
  - r/DungeonsAndDragons
  - r/DataIsBeautiful
  - r/Gephi



# The Reddit Comment Database

**dataset** | I have every publicly available Reddit comment for research. ~ 1.7 billion comments @ 250 GB compressed. Any interest in this? self.datasets  
Submitted 2 years ago \* by Stuck\_In\_the\_Matrix - [pushshift.io](#)  x2

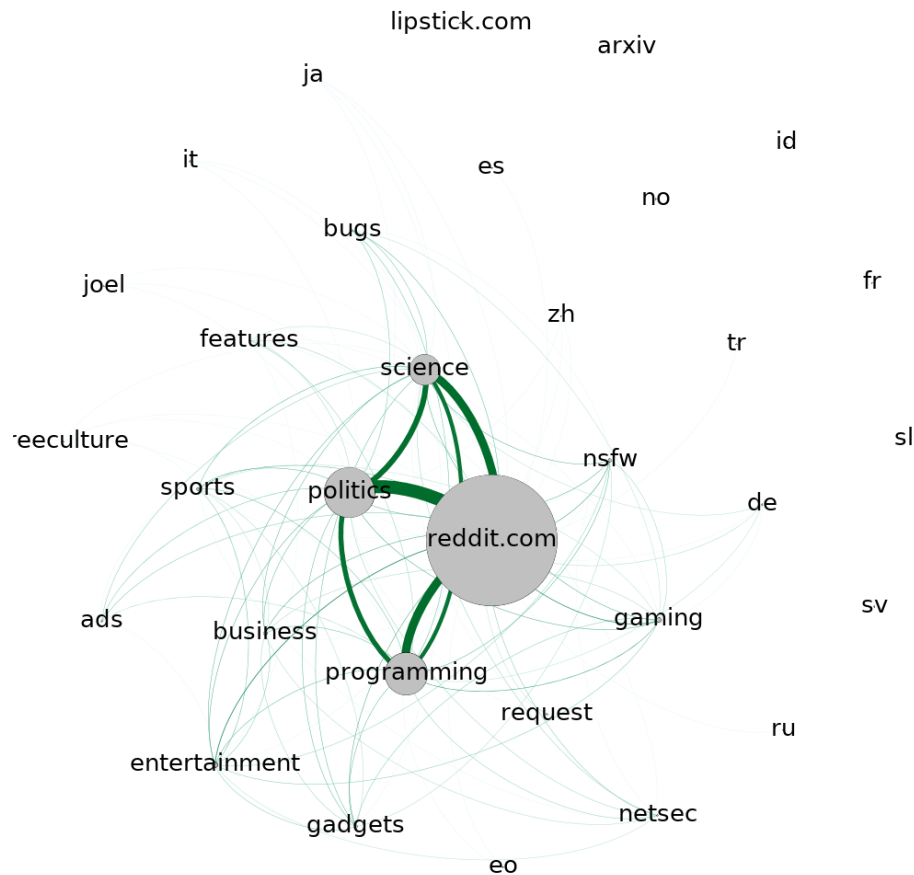
- **Complete dataset of reddit comments, divided by month, since October 2007**
  - Author name
  - Body of comment
  - Score
  - Thread
  - Subreddit
  - Timestamp
  - ...

# The Subreddit Graph

- **Nodes represent a subreddit, with attributes:**
  - Number of comments
  - Number of unique comment authors
- **Edges connect subreddits:**
  - Undirected connection
  - Edge weight corresponds to the number of shared authors between the communities

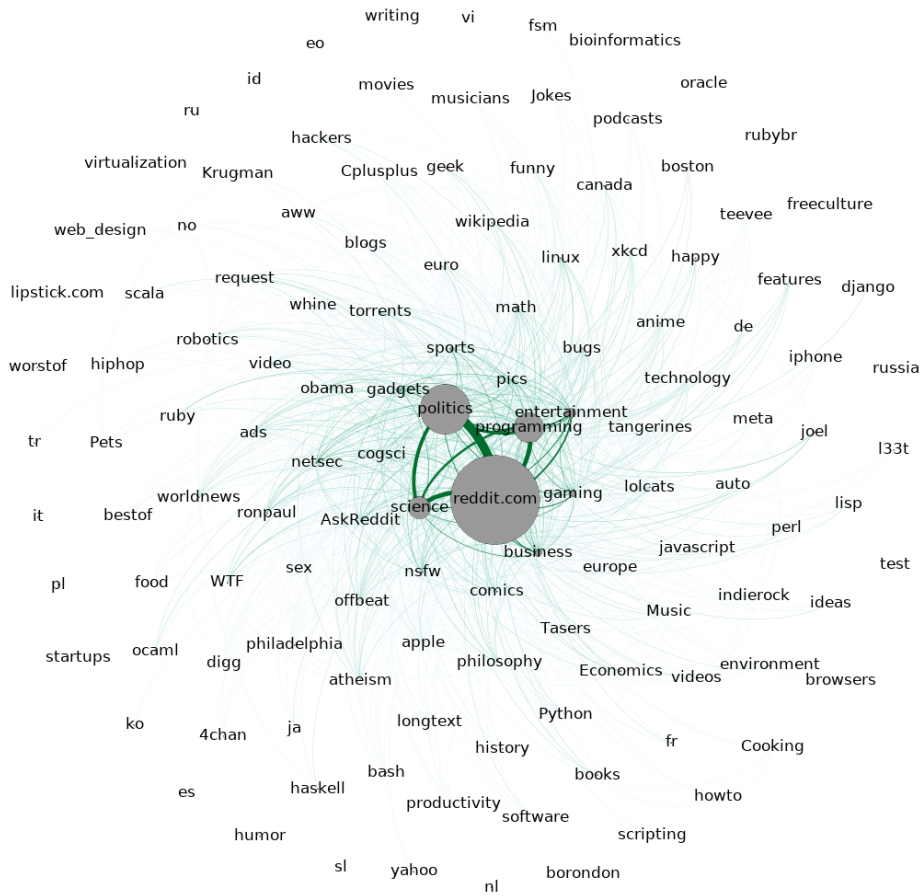
**Comment processing was done with Python**  
**Graph analysis was done with Gephi**

# Subreddit Graph for October 2007



- **32 subreddits**
- **104 connections**
- **The main page of reddit was a separate subreddit back then. Today it's a collection of links from all subreddits.**

# Subreddit Graph for January 2008

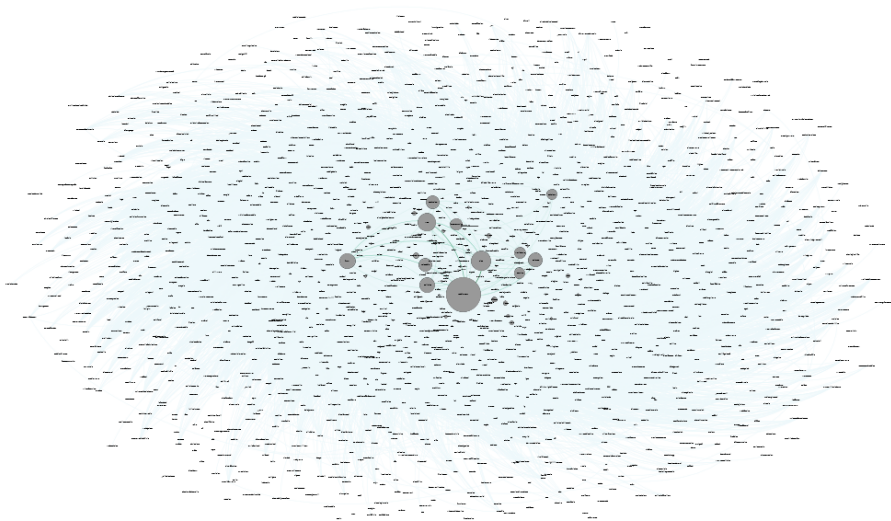


- **126 subreddits**
- **856 connections**
- **Boom of subreddits, but still 4 central:**
  - Reddit.com
  - Politics
  - Science
  - Programming



# Subreddit Graph for July 2009

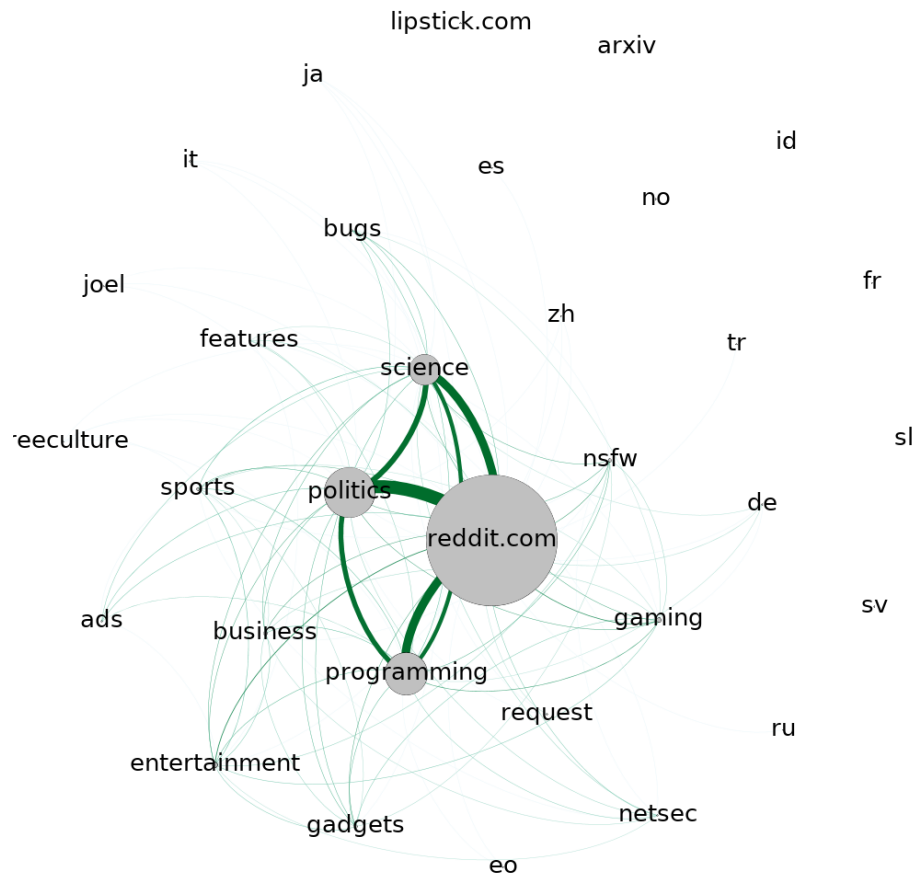
- **1686 subreddits**
- **60733 connections**
- **A dozen subreddits start to gain more popularity**



# **Analysis of Social and Information Networks**

## **Chapter 3 – Strong and Weak Ties**

# Clustering Coefficient Over Time



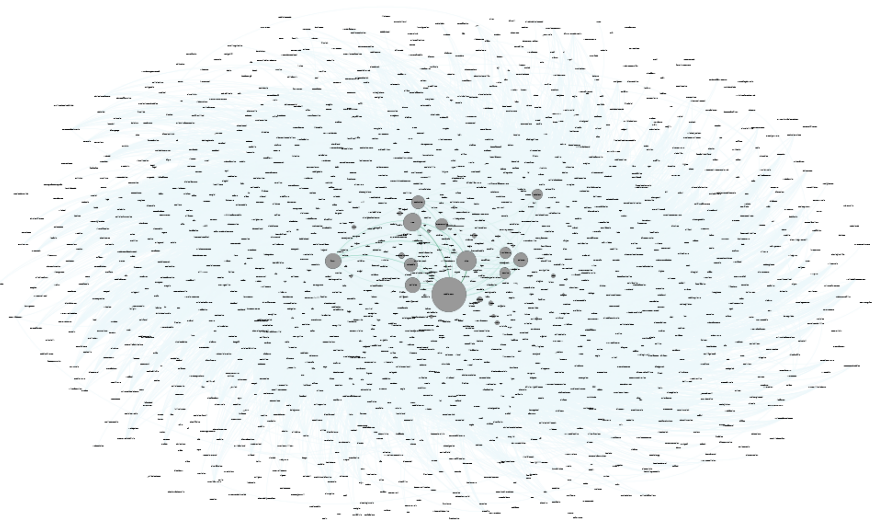
- **Users tend to gravitate towards the bigger subs**
  - Lots of shared users between all subreddits
  - Isolated subreddits are usually very small
  - High number of connections between nodes
  - High Clustering Coefficient  $\approx 85\%$

# Clustering Coefficient Over Time

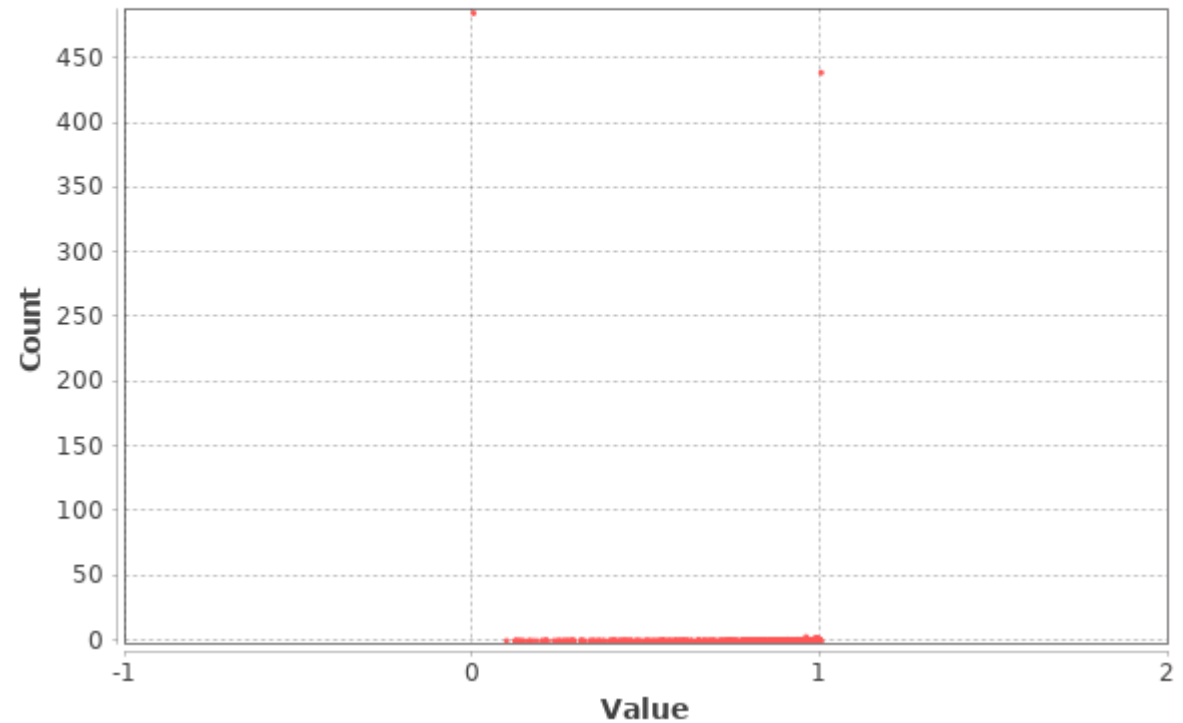
Average Clustering Coefficient: 0.875

Total triangles: 2463055

The Average Clustering Coefficient is the mean value of individual coefficients.



**Clustering Coefficient Distribution**



# Clustering Coefficient Over Time

- **Small increase overall of the Clustering Coeff.**

- All subs tend to be connected, considering it just takes one shared author to establish a connection
- Low number of local and global bridges
- Increase in coeff. can be due to growing subs with more diverse commentators: higher participation in a large number of subs will increase the number of connections between all

- **Triadic closure usually exists:**

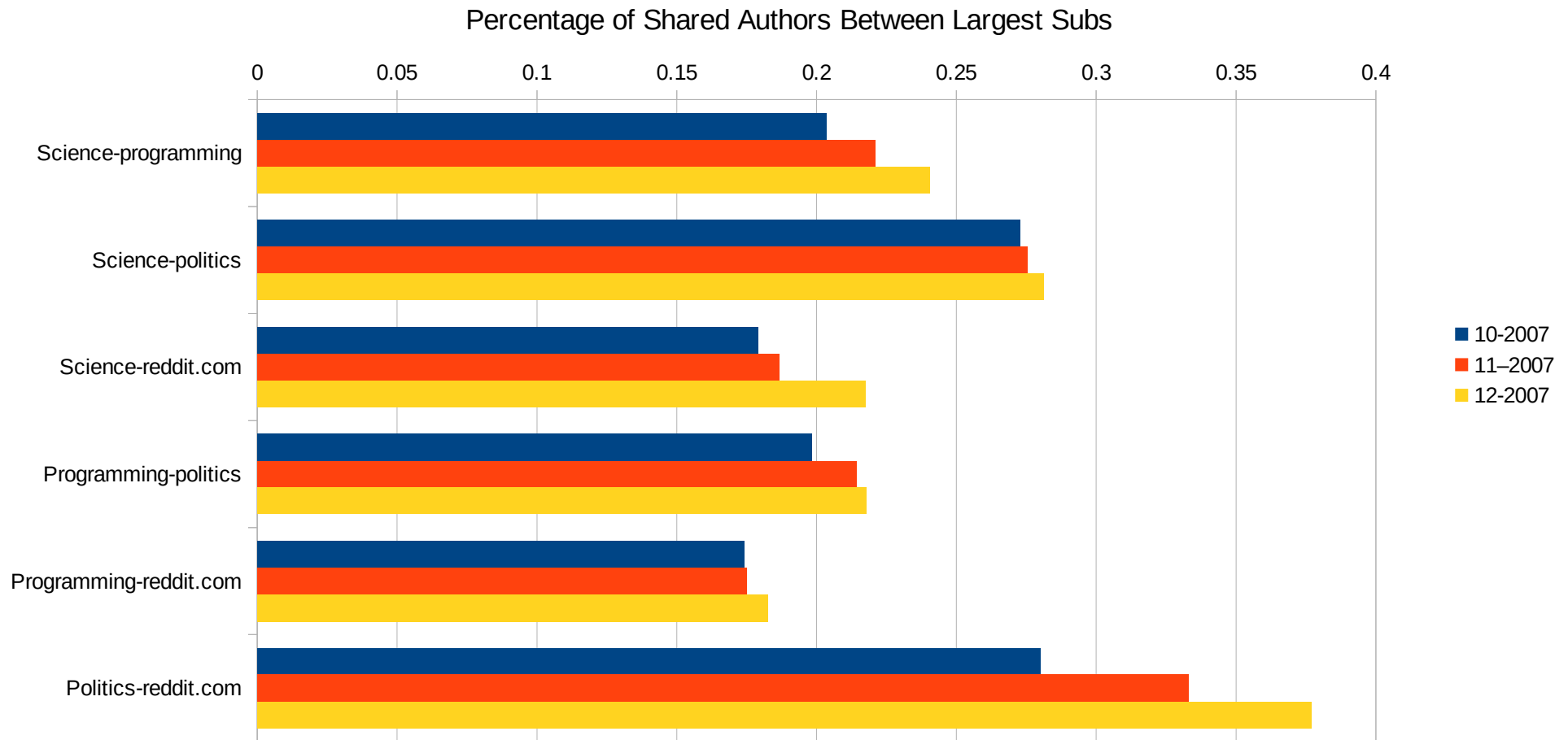
- If 3 subs share one author, all of them will be connected

- **More interesting to analyze the evolution of shared authors between specific subs**

- “**Diadic Reinforcement**”: subs that are connected will tend to increase the weights between them

# Diadic Reinforcement Between Subs

**Connected subs become increasingly more connected**



# Diadic Reinforcement Between Subs

- **It's hard to say if there is a real effect of Triadic Closure happening**
  - A connection to another sub probably doesn't affect its users to actively engage in the other, unlike with friend networks
  - Since these are the 4 most popular subs at the time, it's more likely that more users visit and interact with them, since they are growing in popularity
- **The increasing popularity of the site brings new users to all the subs, increasing their connections**

# **Analysis of Social and Information Networks**

## Chapter 4 - Networks in Their Surrounding Contexts



# Homophily and Reddit

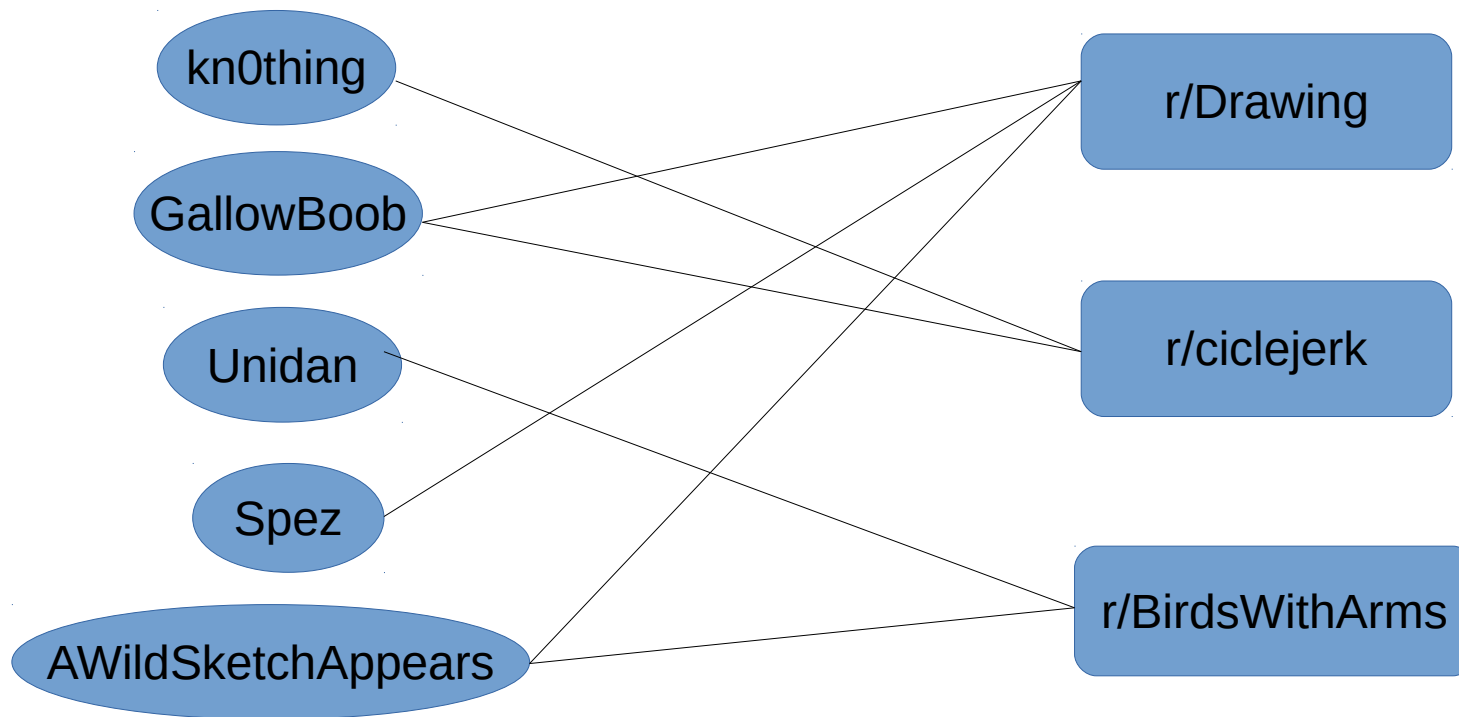
- **Reddit.com presents itself as a main hub for a plethora of topics, interests and information**
- **The comment authors are anonymous: no information is known about them other than their username**
- **However, the concept of a subreddit puts in homophily in motion:**
  - The creation of specific communities for specific topics will bring together users who share interest in these topics
  - Subreddits create more uniform user distributions in them
  - Still very diverse: the Internet is a very large place, with users from everywhere and with every possible interest

# Selection and Social Influence

- **Subreddits are a perfect example of user Selection**
  - A user can select which subreddits it wants to visit and subscribe to, thus creating a personalized social environment surrounded by users with similar tastes and interests
- **But Social Influence also plays a big roll on Reddit**
  - Reddit Hive Mind Thinking: overall, users tend to like and hate the same things
  - Scoring system only benefits popular opinions, which in turn get more popular; provides direct feedback of the popularity of a personal opinion
  - “Circlejerking”: A group discussion or activity between like-minded individuals that validates mutual biases or goals in a non-confrontational environment.

# Affiliation in Subreddits

**We can consider a graph of comment authors affiliated with the subreddits they post on**



# Closures in the Subreddit Context

## • **Triadic Closure:**

- Unlikely to happen, since user interaction tends to be random, with no specific bias for selected authors

## • **Memembership Closure:**

- Unlikely, a interaction with a certain user will not create a tendency for it to visit an affiliated subreddit
- Unless said interaction is in a affiliated sub, in that case the closure is automatic

## • **Focal Closure**

- If two users are affiliated with a specific subreddit then the probability of interaction between them increases.
- This considering said interaction is intra-sub; inter-sub there is most likely no influence

# **Analysis of Social and Information Networks**

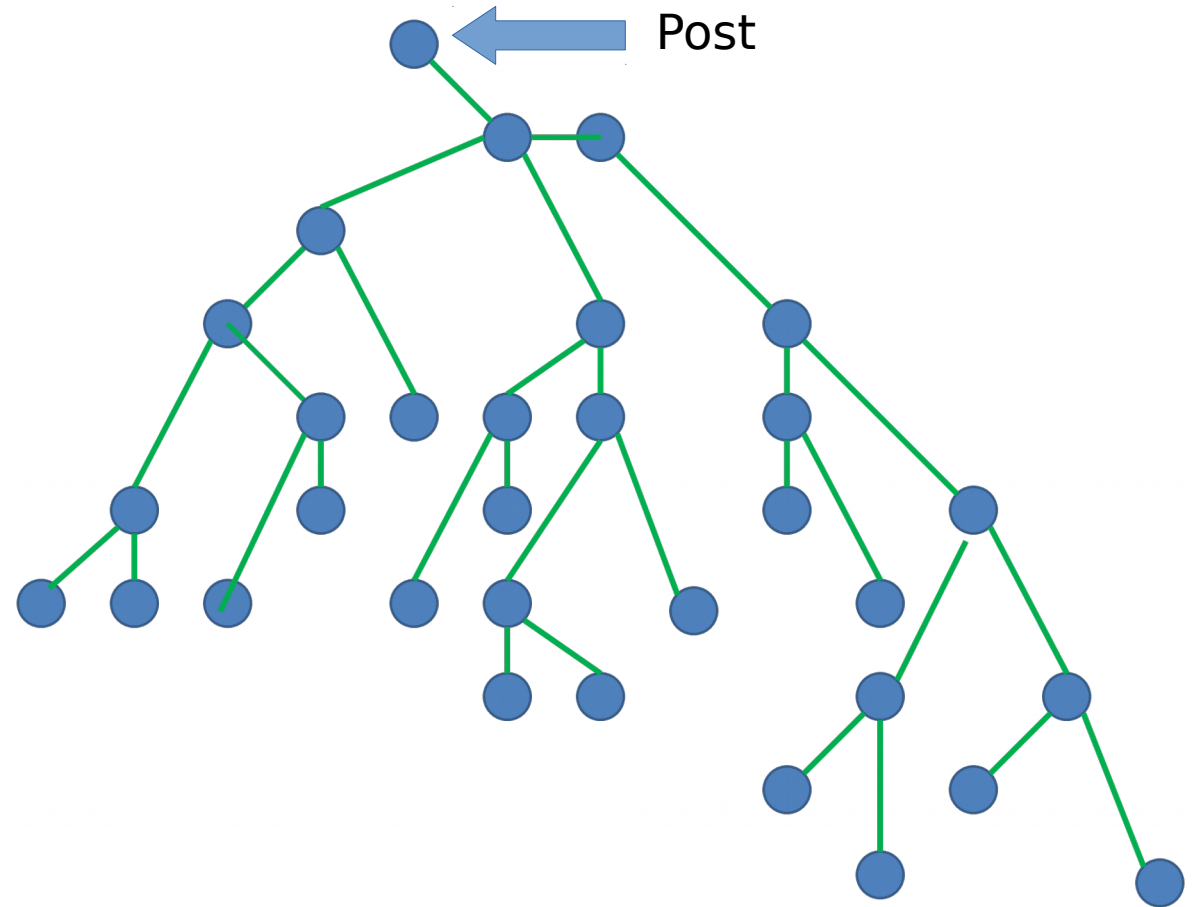
## **Chapter 5 – Positive and Negative Relations**

# Relations in Reddit

- **Users can mainly interact in two ways:**
  - Comment on a link
  - Respond to a user comment on a link
- **Reddit's scoring system provides insight to a comment/link content:**
  - Upvoted comments/posts tend to present common and accepted opinions by the community
  - Downvoted comments/posts pose unpopular opinions or irrelevant content for the discussion (e.g. spam)

# Graphs of Comments

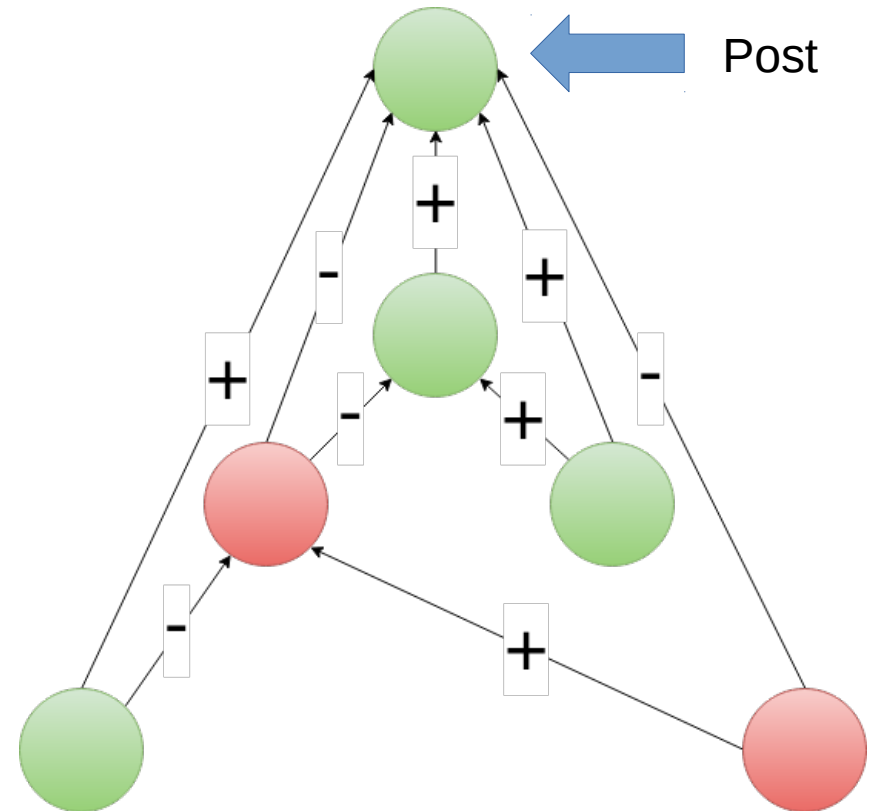
- **Consider that a comment is a response to a post or another comment**
  - Creates graph trees of responses
  - One comment can only respond to another comment, but one comment can originate several responses
  - No Triangles



# Graphs of Comments: Another Approach

- **Consider each comment is a response to the post AND another comment**

- Generates 1 triangle per response
- Green: upvoted
- Red: downvoted
- Agreeing opinions: +
- Disagreeing opinions: -
- Balanced





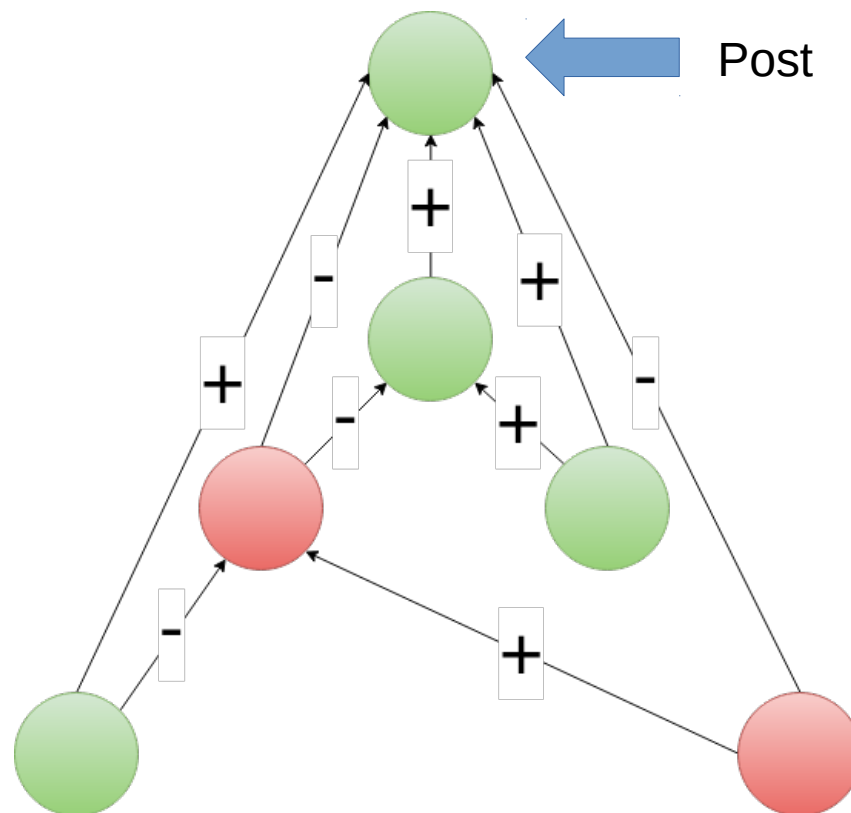
# Graphs of Comments: Assumptions and Challenges

- **Assumptions:**

- The score of a comment/link is a measure of how agreeable is its content.
- If a response has the same score as its parent, they are of the same opinion
- If a response has a symmetric score as its parent, they are in disagreement

- **Interesting to prove these assumptions:**

- Text mining to determine sentiment and level of agreement of a comment/post
- Determine if a post/subreddit has balanced comments
- Determine opposing communities on a subreddit (e.g. r/politics before a major election)



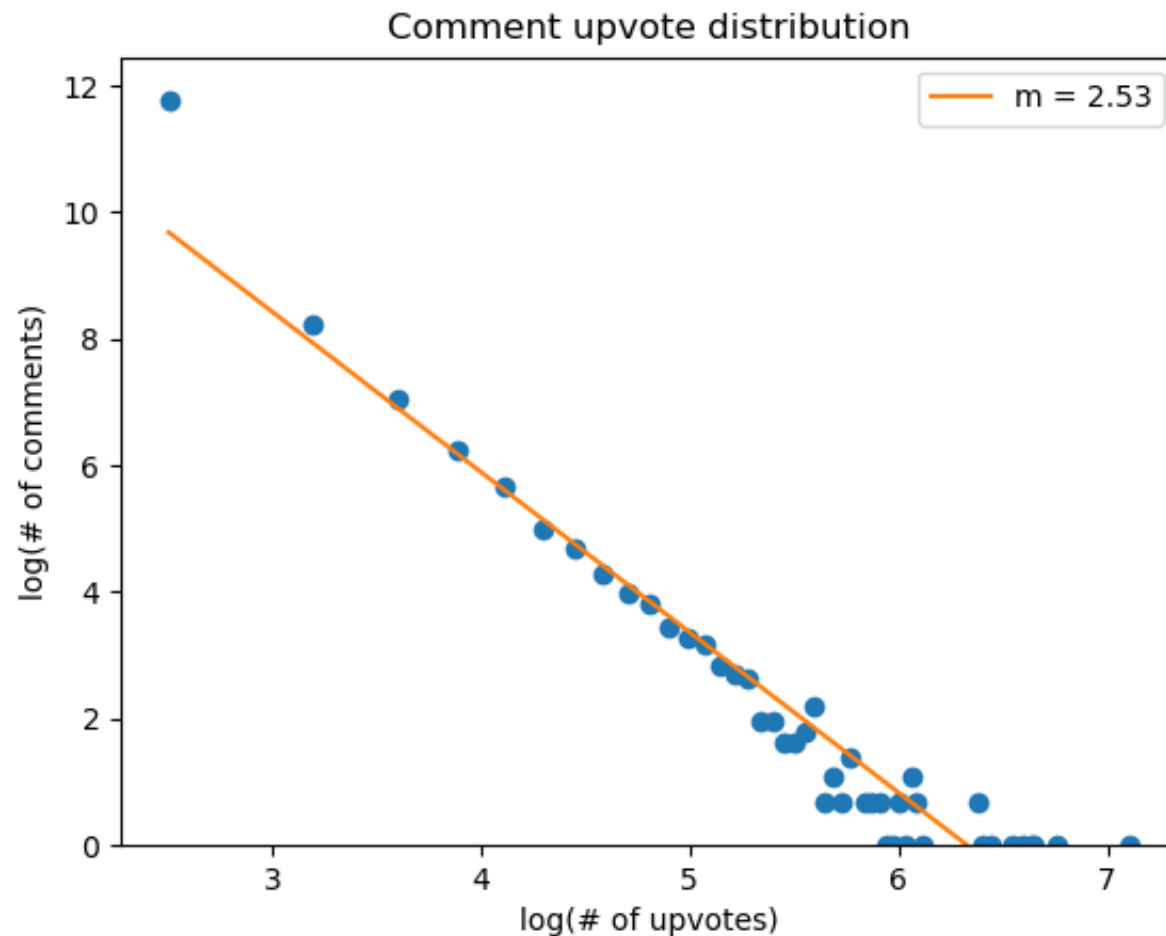
# **Analysis of Social and Information Networks**

Chapter 18 - Power Laws And Rich-Get-Richer Phenomena

# Popularity on Reddit

- **Number of upvotes determines popularity of content**
- **Rich-get-richer phenomena:**
  - Upvoted content is shown first to any user, causing it to become more upvoted and popular
- **Popularity and power laws:**
  - Several distributions will present a power law model for popularity:
    - Comments and upvotes
    - Links and upvotes
    - Links and number of comments

# Comments and number of upvotes



Distribution of comments and respective upvotes,  
calculated using a 100-bin histogram

# **Analysis of Social and Information Networks**

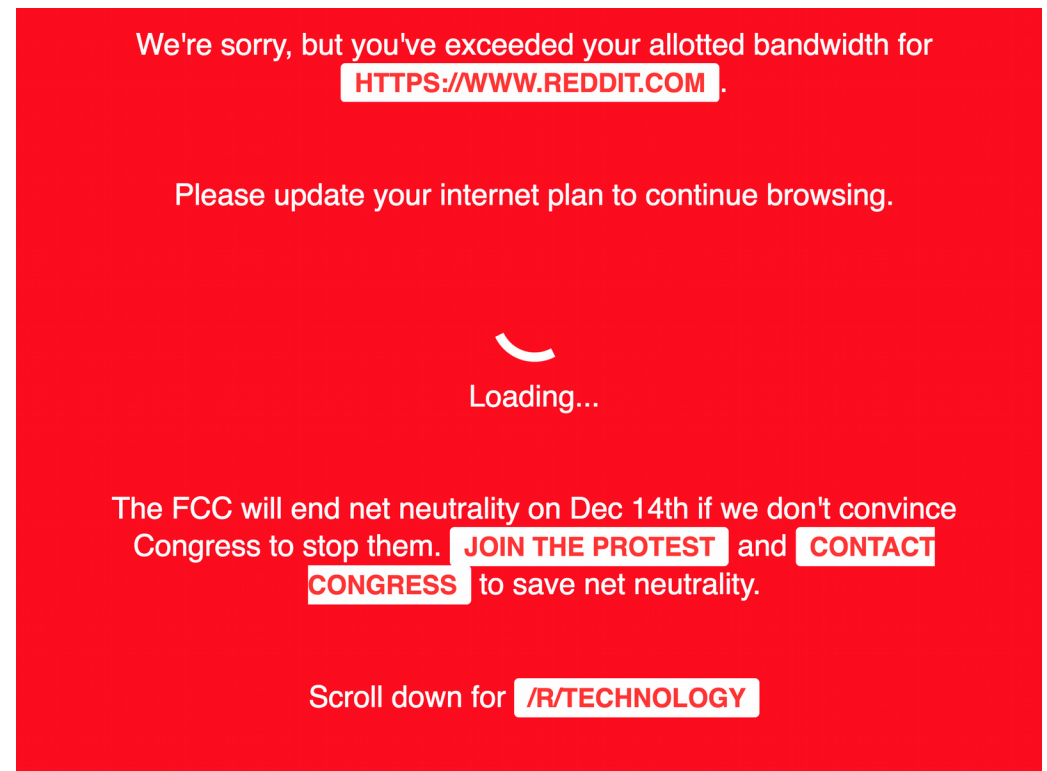
Chapter 19 – Cascading Behavior in Networks

# Diffusion of Opinion on Reddit

- **Reddit has very homophilic communities**
  - Subreddits are created for one topic/purpose = Isolated Clusters
  - Controversial/unpopular opinions are usually downvoted: less impact on the community
- **Adoption/Change of opinion is very unlikely on a large subreddit**
  - Subreddits tend to share the same general opinion: Reddit Hive Mind, Circlejerking
  - If a group of users does not feel represented in a subreddit, they create their own for that topic

# Collective Action on Reddit

- **The power of anonymity decreases the risk of repercussions from collective actions**
- Some events unite all the subreddits with a common goal



# r/politics and Political Support

- **r/politics is a subreddit for political discussion**
  - Usually U.S.A. politics
- **On 2008, prior to the Obama vs. McCain Election:**
  - Most of the links and comments supported Obama
  - Only a few unpopular submissions were supporting McCain

If idiot conservatives will argue that Obama=Osama, then I submit that Palin=Stalin. (self.politics)  
submitted 17 hours ago by Ostrich159  
420 comments share

NEW SCANDAL??? It's possible that the contractor that built Palin's ill-advised, overpriced sports complex built her \$500,000 house - for free! (stopthinkvote.com)  
submitted 11 hours ago by beenieseigleman  
55 comments share

John Cleese on Sarah Palin (youtube.com)  
submitted 3 hours ago by saki  
6 comments share watch

Rachel Maddow on Calling a liar, a liar (youtube.com)  
submitted 8 hours ago by IAmPerfectlyCalm  
10 comments share watch

Disgraceful: 70% of military absentee ballots will not be counted (washingtontimes.com)  
submitted 12 hours ago by r2002  
34 comments share

Maddow chats with Frum about tone in the election. Frum tries to equate Maddow's show with threats of violence. So Maddow kicks his ass. (youtube.com)  
submitted 9 hours ago by obscureant  
57 comments share watch

Obama's a Muslim? Meet the man who started that false rumor ... a man so mentally unbalanced he wasn't allowed to become a lawyer, but Fox gives him air time. (nytimes.com)  
submitted 21 hours ago by AMerrickianGirl  
321 comments share



# **Analysis of Social and Information Networks**

Chapter 20 – The Small World Phenomenon

# Social Distance with Reddit Links

- **Reddit links can be compared by the number of commentators they have in common:**
  - The more users two links have in common, the “closer” they are to each other in terms of content or relevancy
- **Distance defined:**
  - The number of shared unique comment authors between links

# Distance Distribution

- **Log-log plot of histogram of distance and counts**
  - Does not have  $p$  close to  $-1$
  - Not efficient
- **But:**
  - Artificial measurement of distance
  - Does not represent social interaction between people

