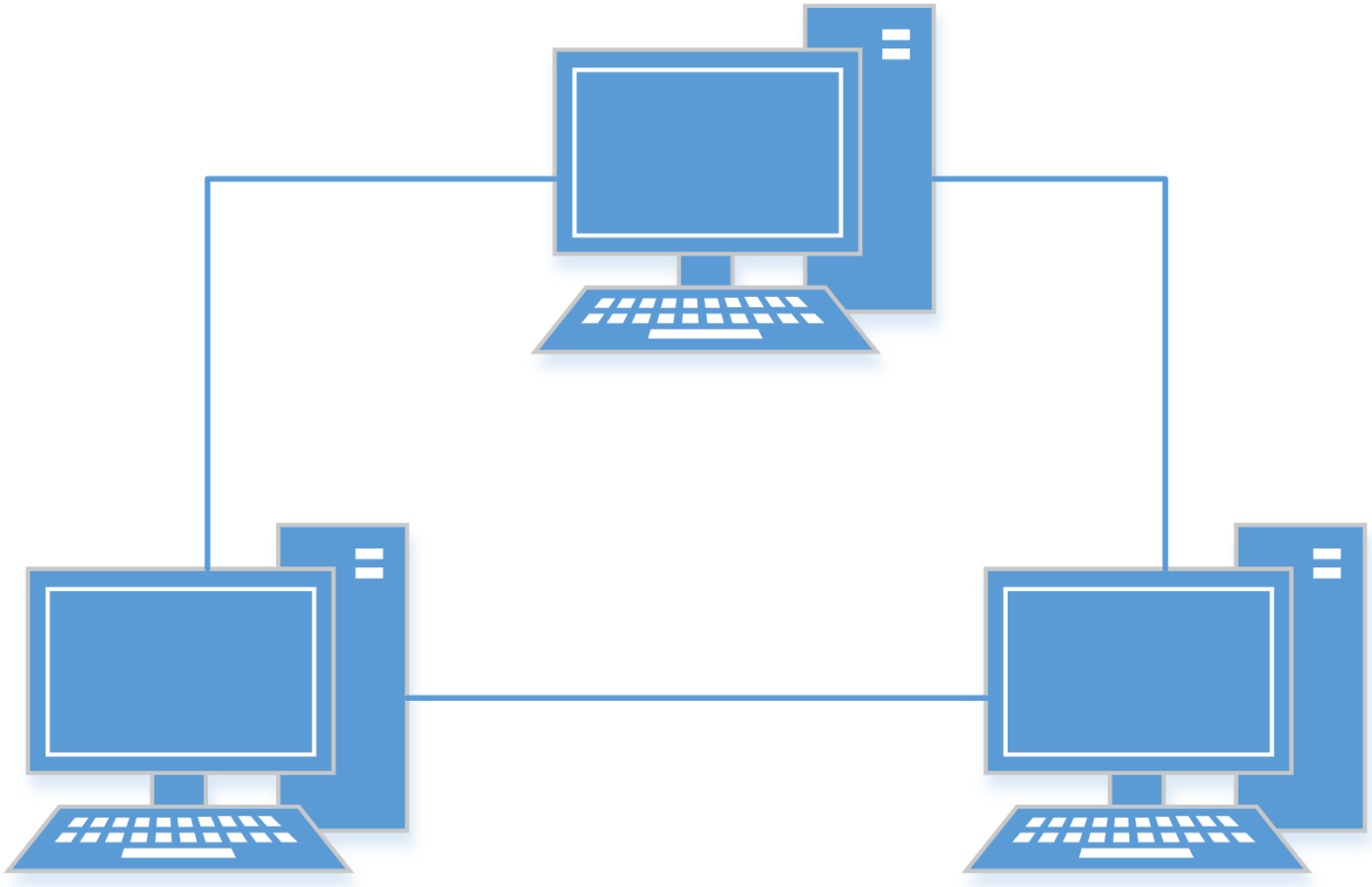
Applying Social Network Analysis (SNA) to P2P File Sharing

Andreas Schaufelbühl

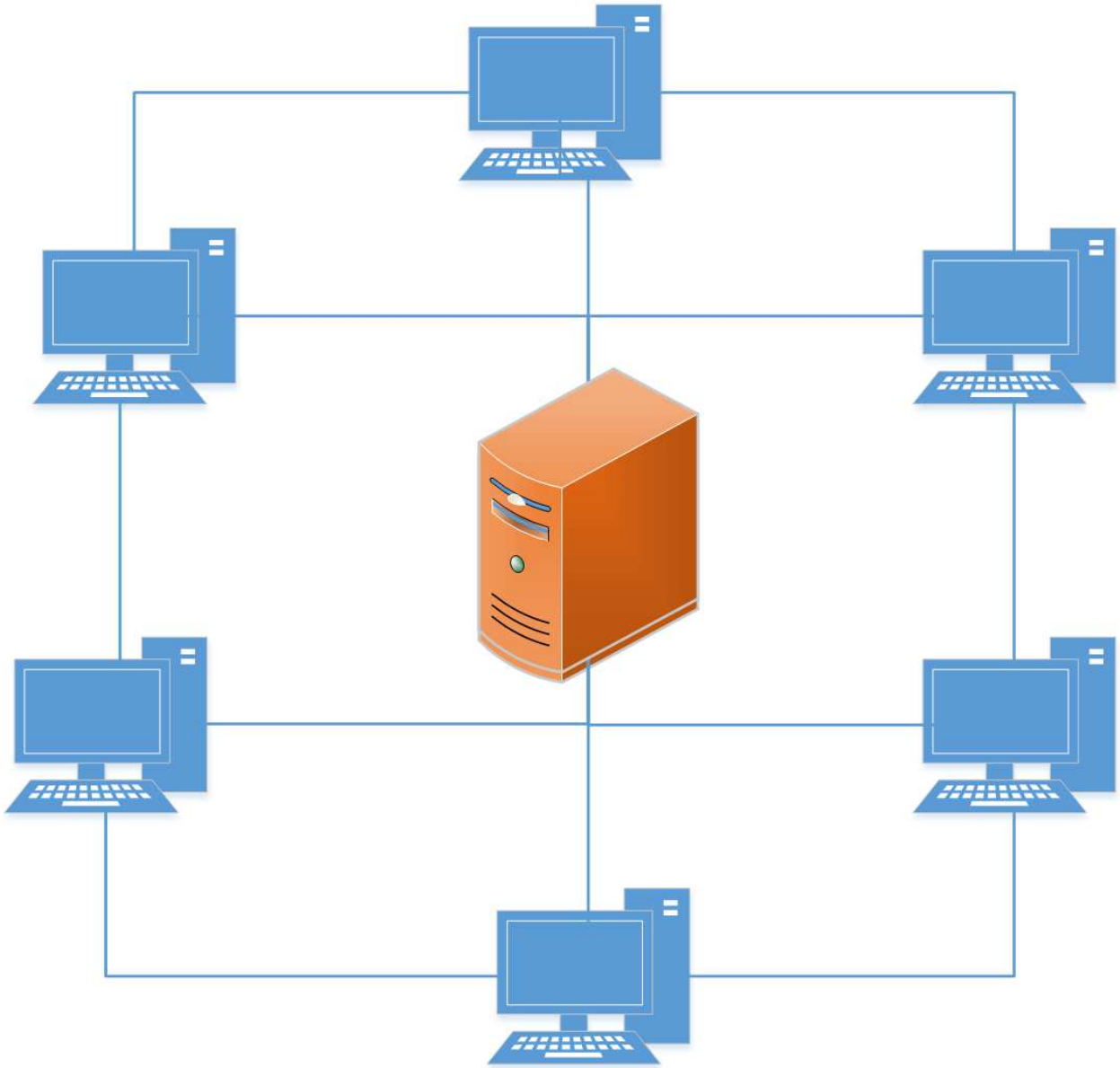
Robin Stohler

Benjamin Bürgisser

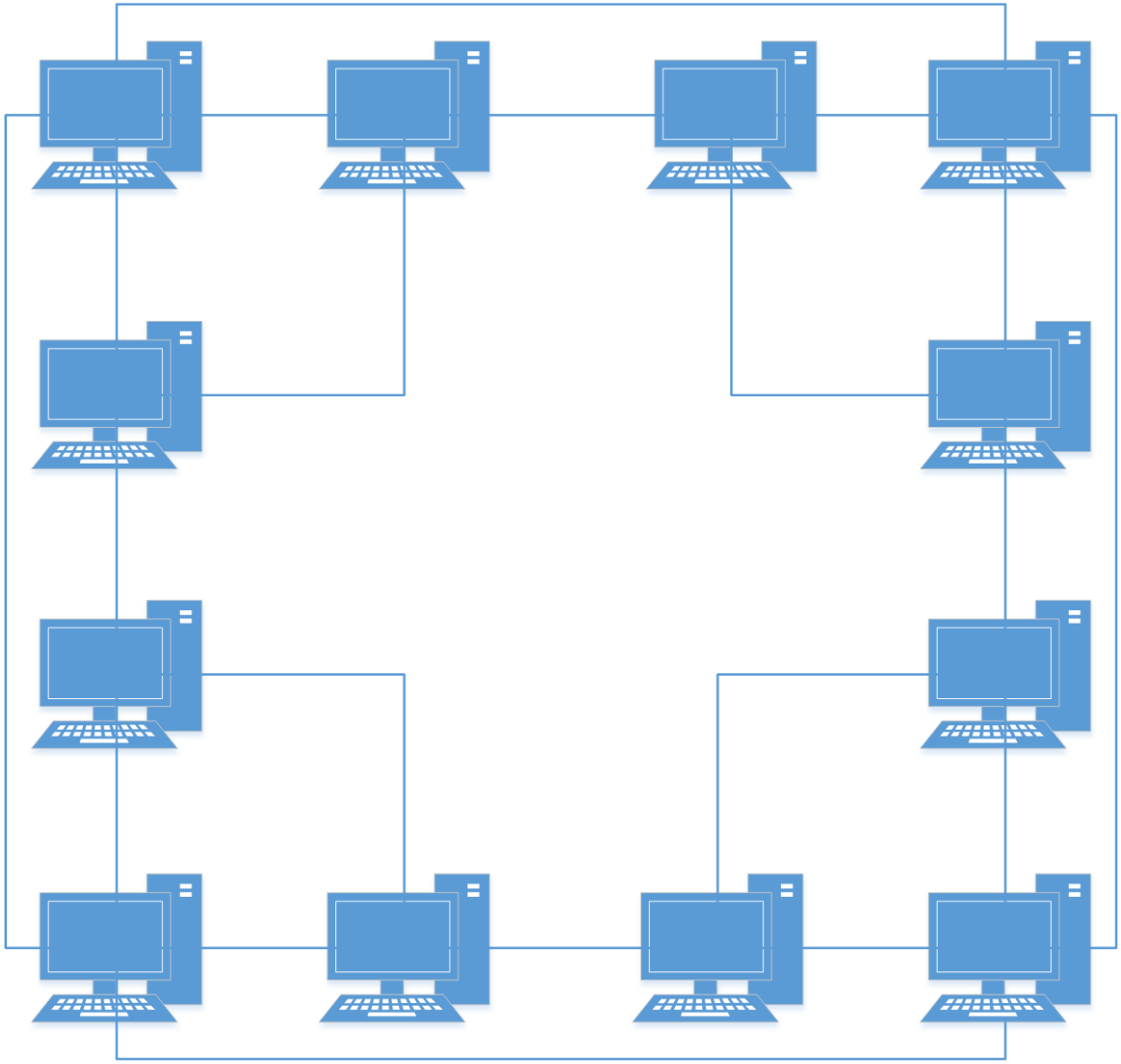
P2P



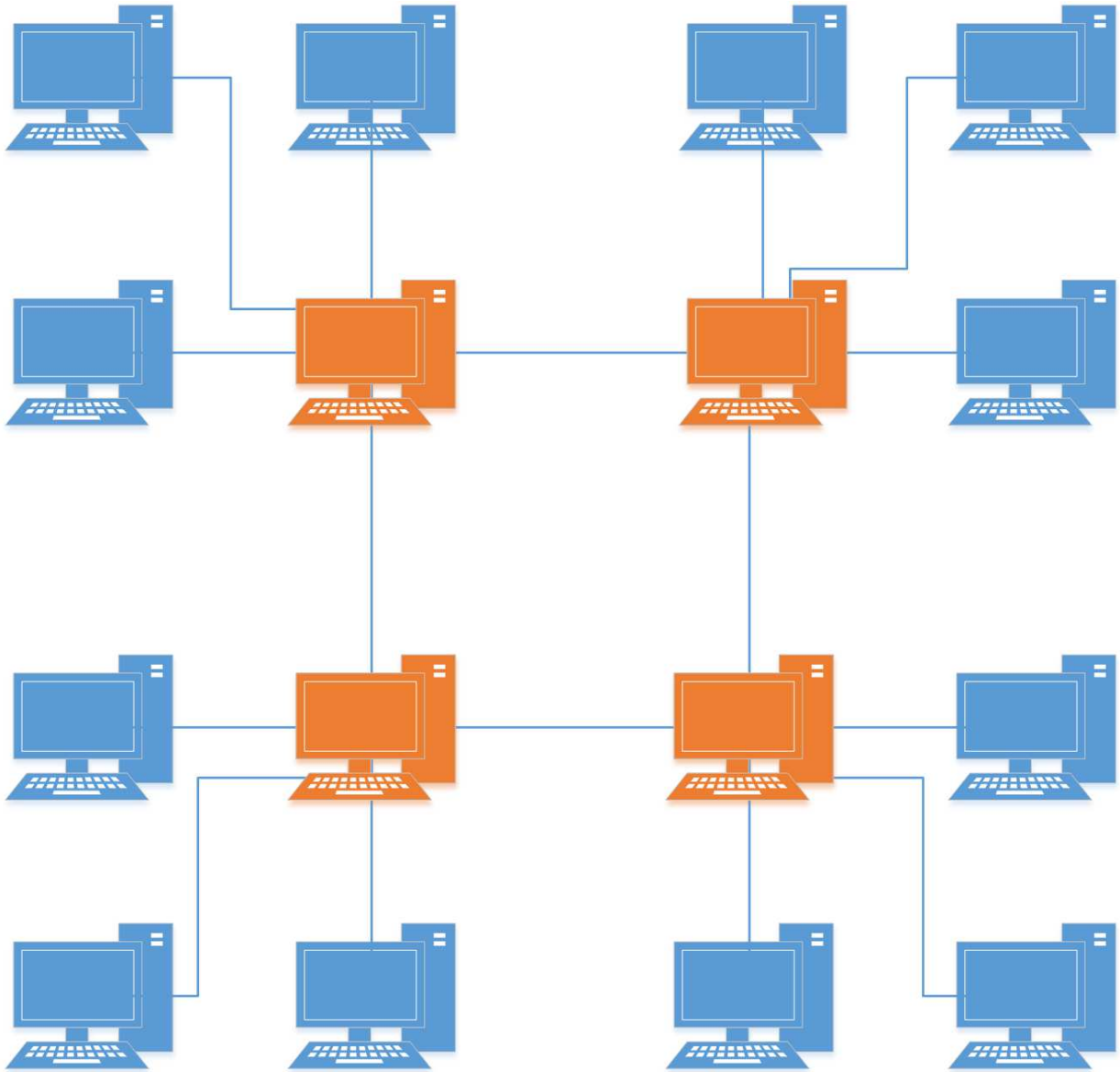
Centralized - Napster



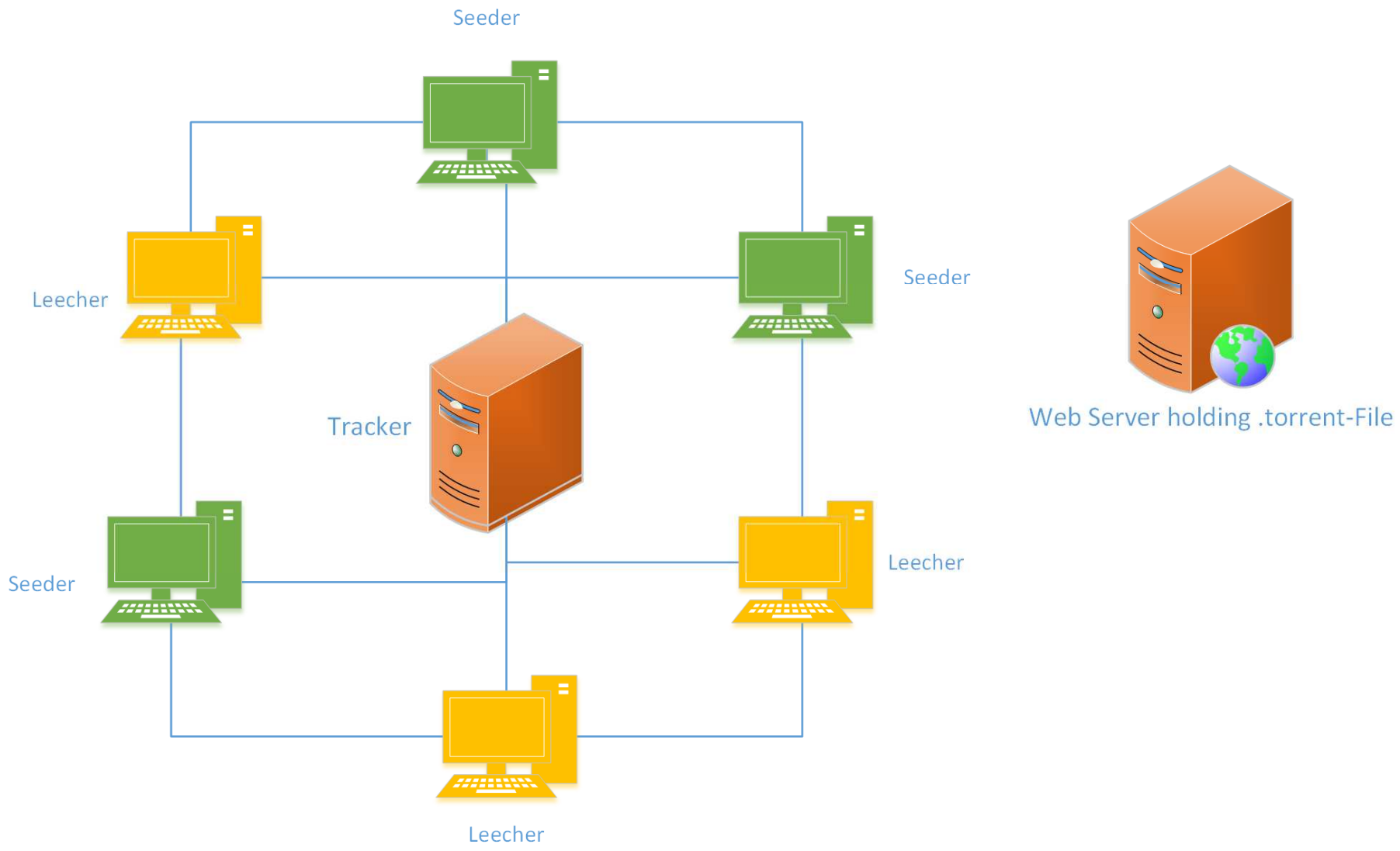
Decentralized - Gnutella 0.4 /Freenet



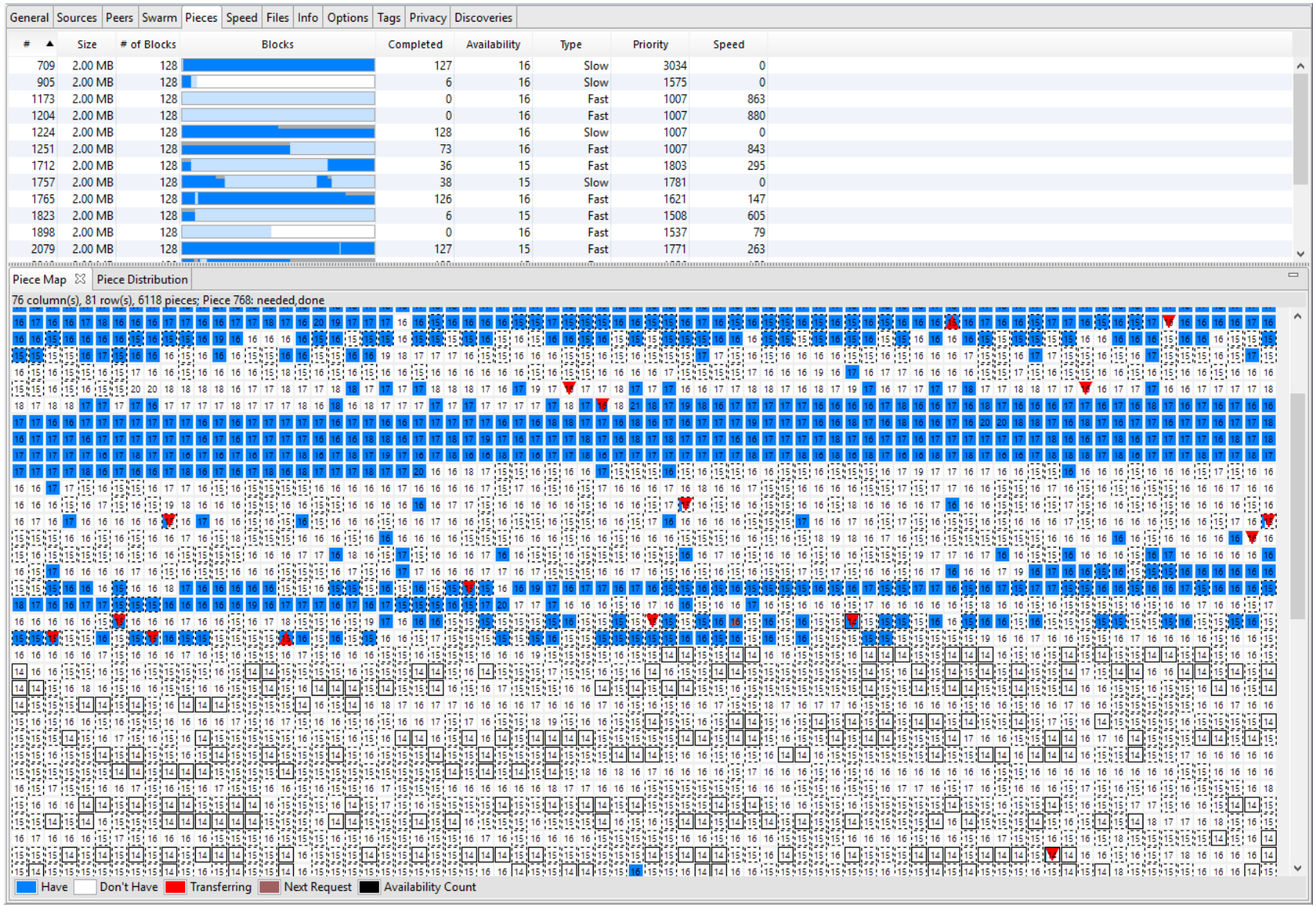
Hybrid - Gnutella 0.6



BitTorrent

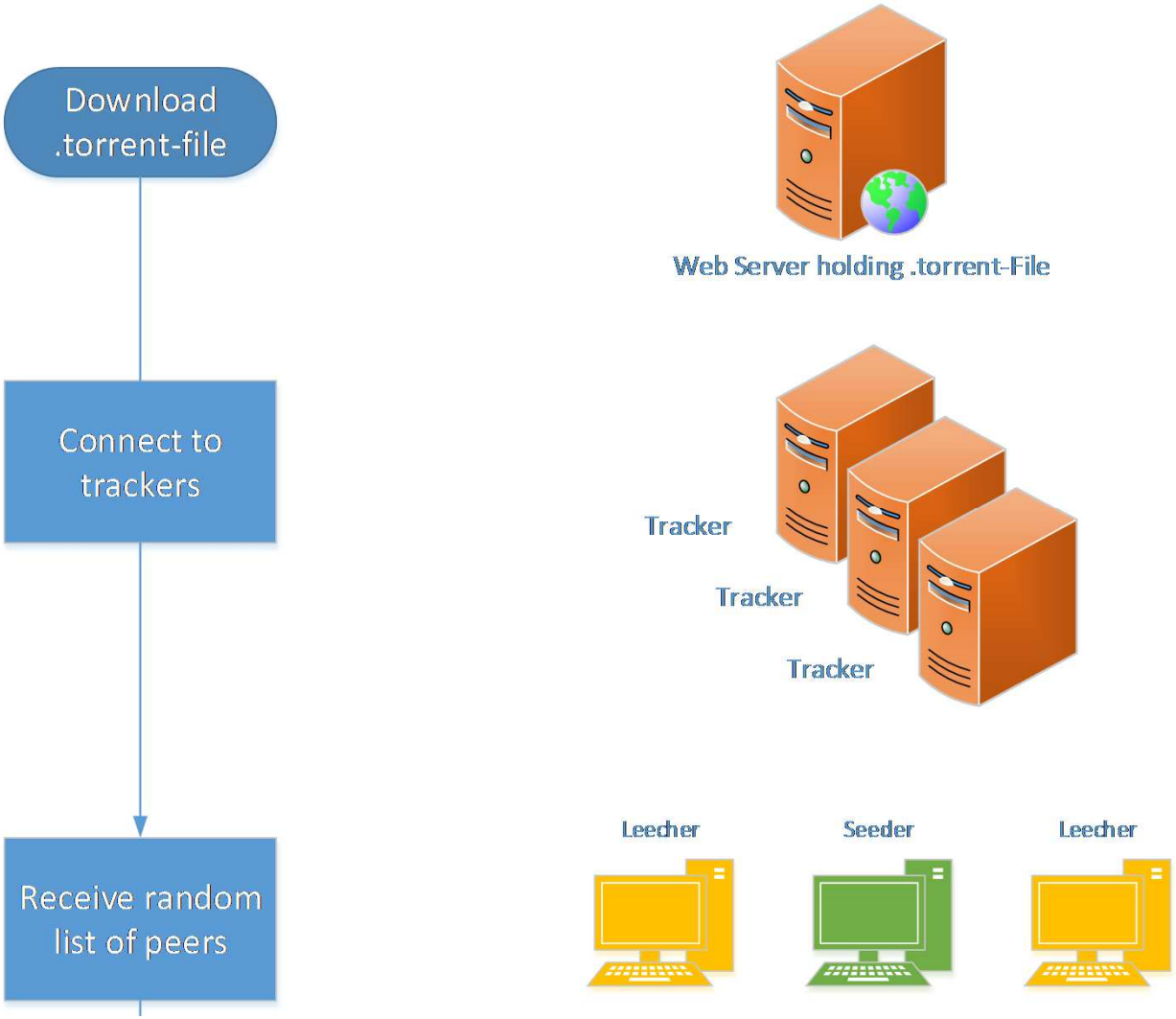

Seeder
Leecher
Seeder
Tracker
Web Server holding .torrent-File
Seeder
Leecher
Leecher

BitTorrent

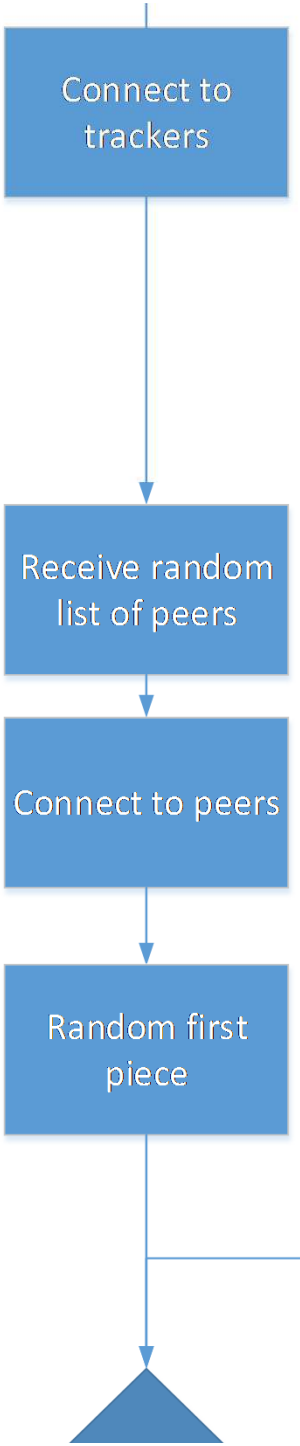


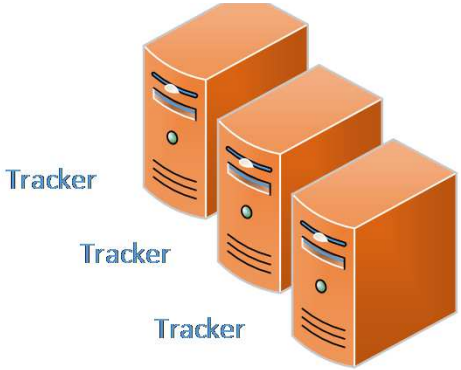
#	Size	# of Blocks	Blocks	Completed	Availability	Type	Priority	Speed
709	2.00 MB	128		127	16	Slow	3034	0
905	2.00 MB	128		6	16	Slow	1575	0
1173	2.00 MB	128		0	16	Fast	1007	863
1204	2.00 MB	128		0	16	Fast	1007	880
1224	2.00 MB	128		128	16	Slow	1007	0
1251	2.00 MB	128		73	16	Fast	1007	843
1712	2.00 MB	128		36	15	Fast	1803	295
1757	2.00 MB	128		38	15	Slow	1781	0
1765	2.00 MB	128		126	16	Fast	1621	147
1823	2.00 MB	128		6	15	Fast	1508	605
1898	2.00 MB	128		0	16	Fast	1537	79
2079	2.00 MB	128		127	15	Fast	1771	263

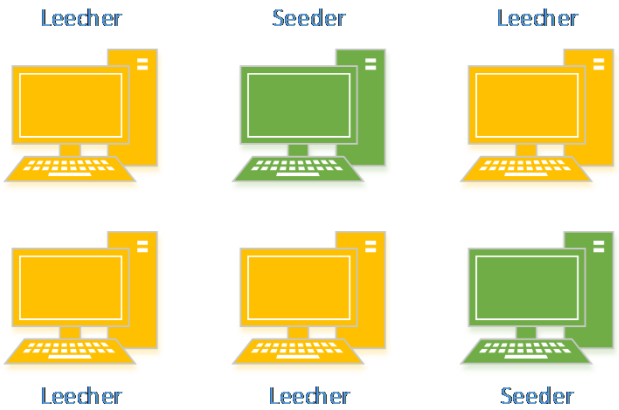
BitTorrent


Download .torrent-file
Connect to trackers
Receive random list of peers
Web Server holding .torrent-File
Tracker
Tracker
Tracker
Leecher
Seeder
Leecher

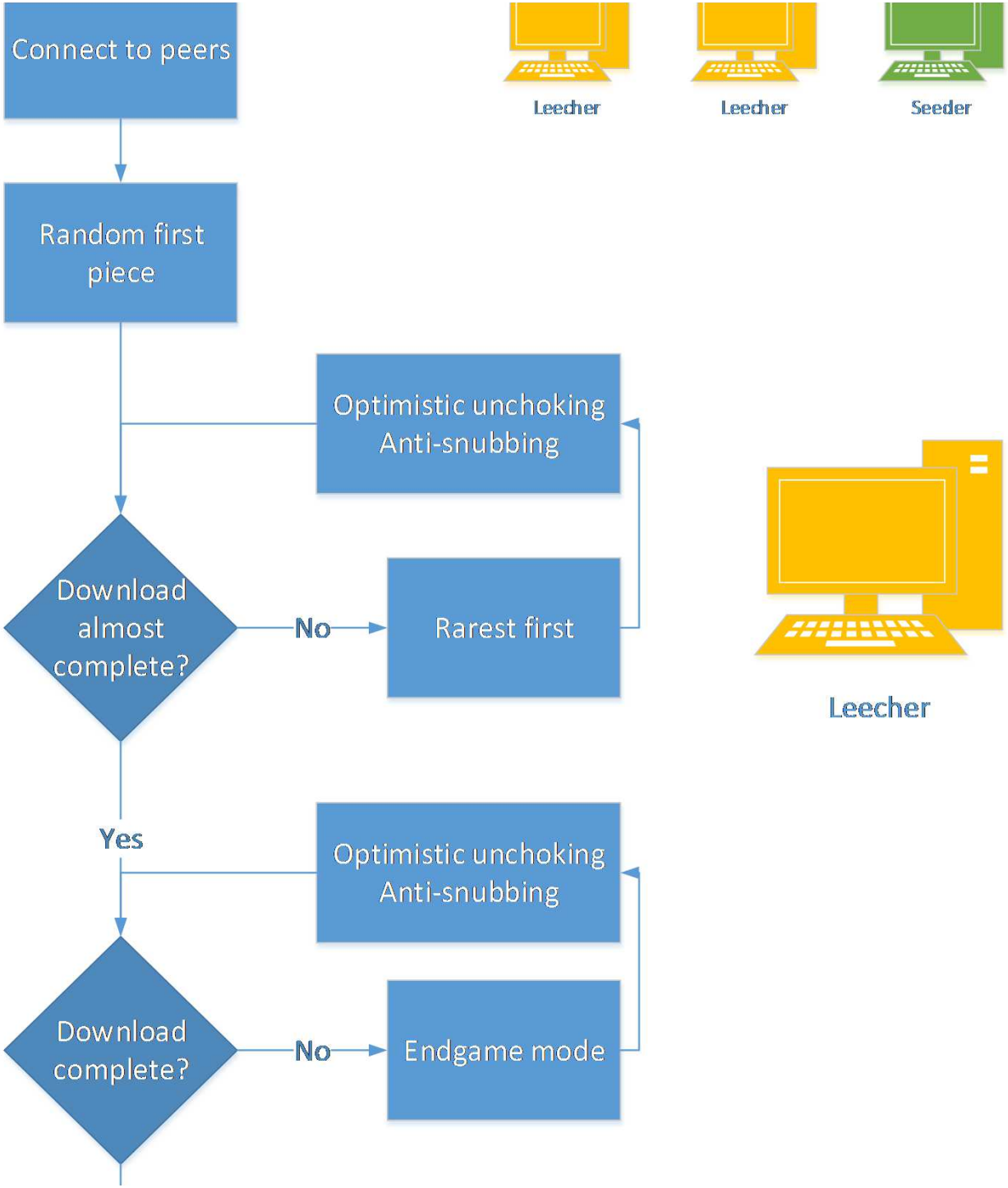
BitTorrent


Connect to trackers
Receive random list of peers
Connect to peers
Random first piece
Optimistic unchoking
Anti-snubbing

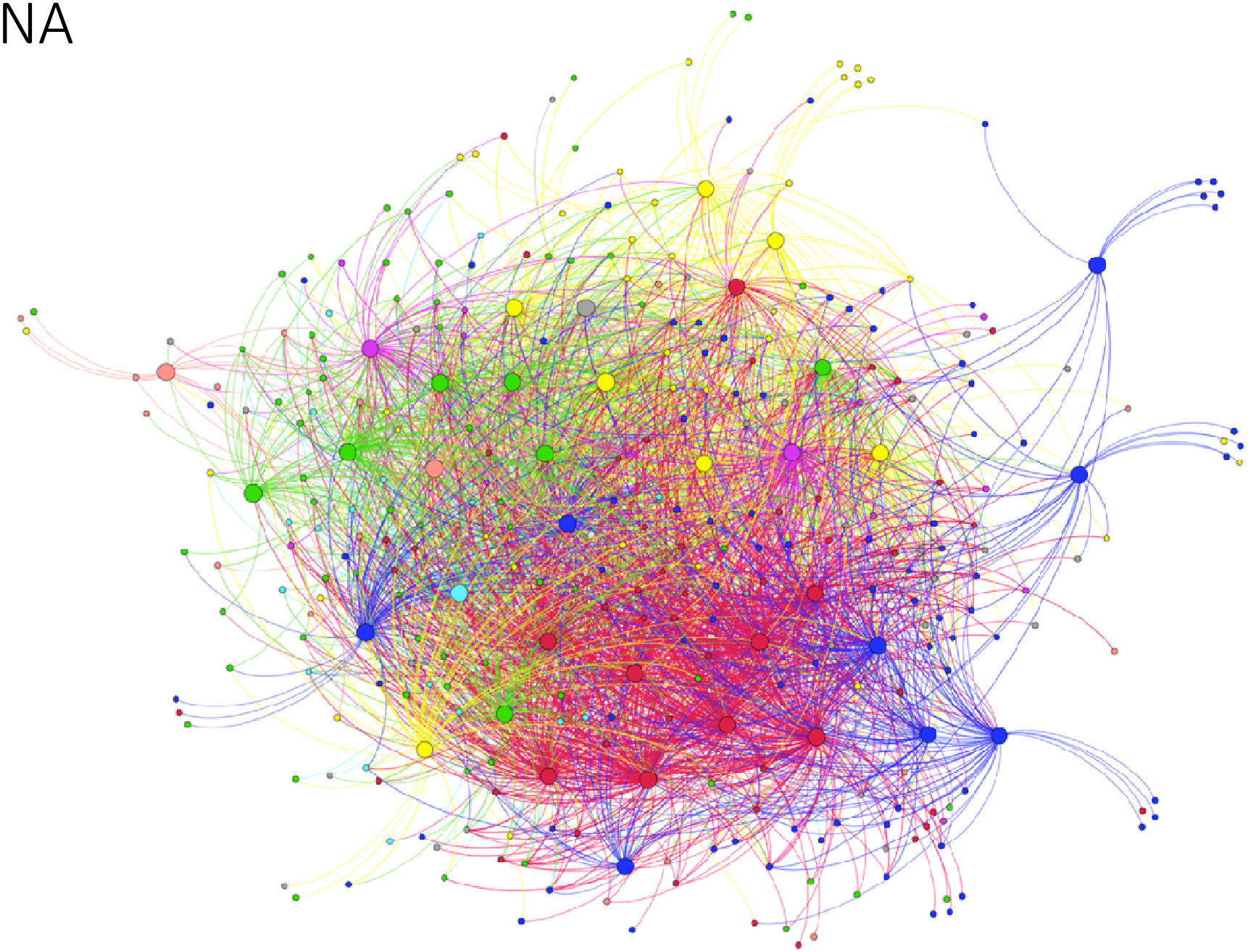

Tracker
Tracker
Tracker


Leecher
Seeder
Leecher
Leecher
Leecher
Seeder

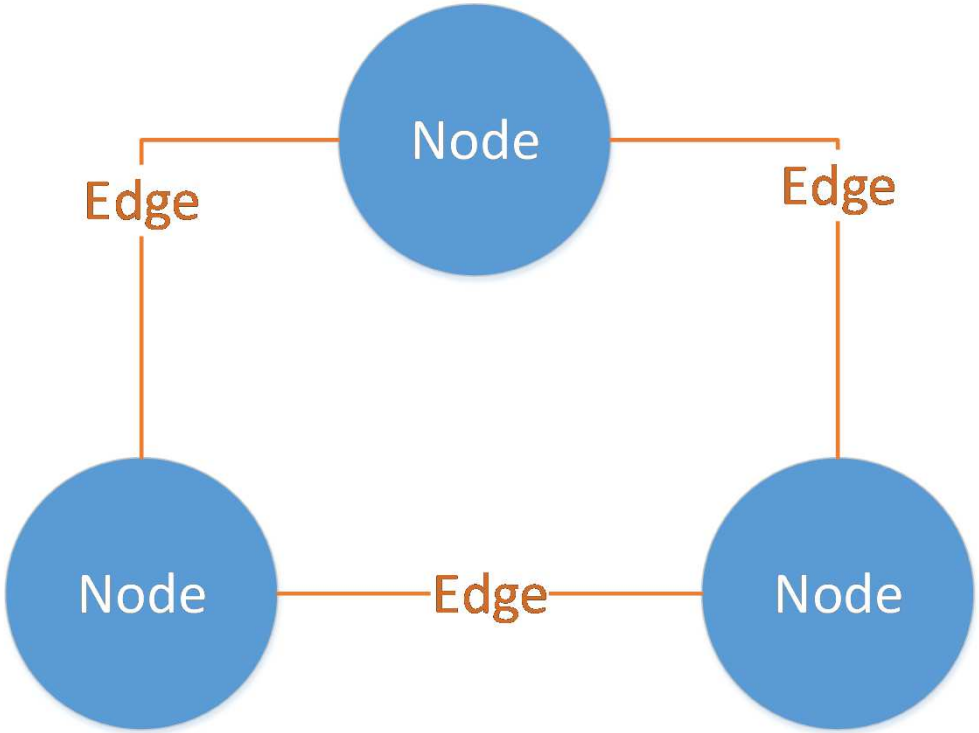
BitTorrent


Connect to peers
Random first piece
Optimistic unchoking
Anti-snubbing
Download almost complete?
No
Rarest first
Yes
Optimistic unchoking
Anti-snubbing
Download complete?
No
Endgame mode
Leecher
Leecher
Seeder
Leecher

SNA



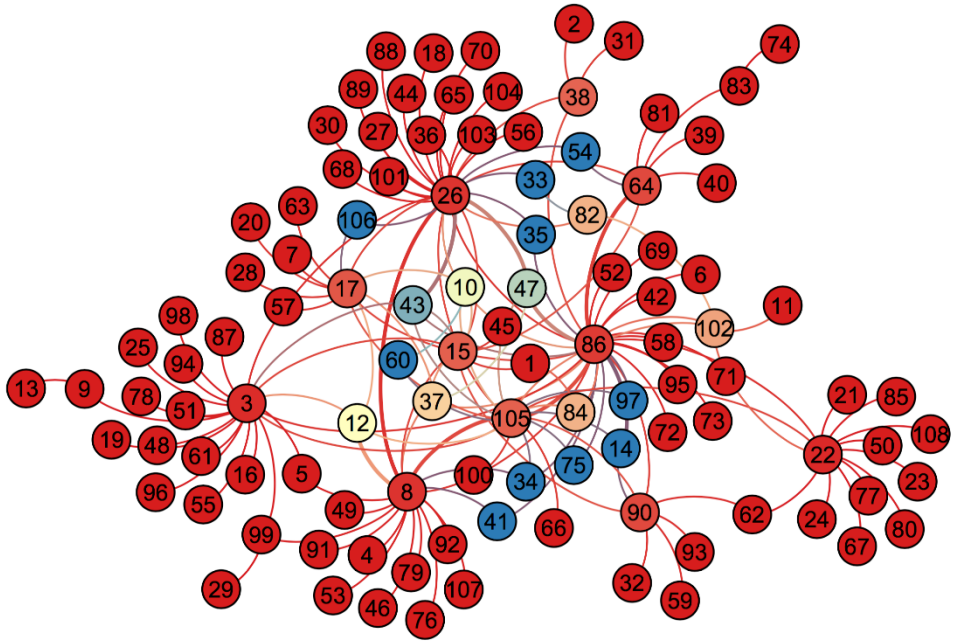
SNA



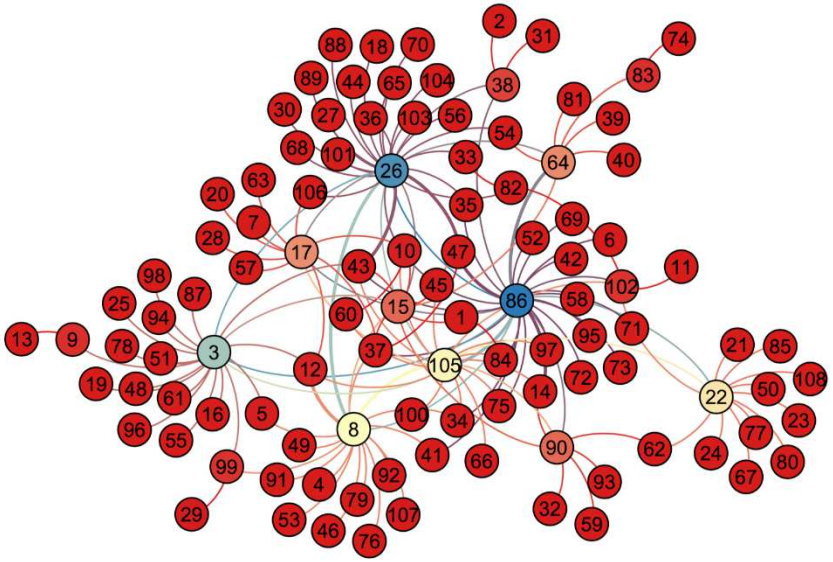
Graph with Nodes and Edges

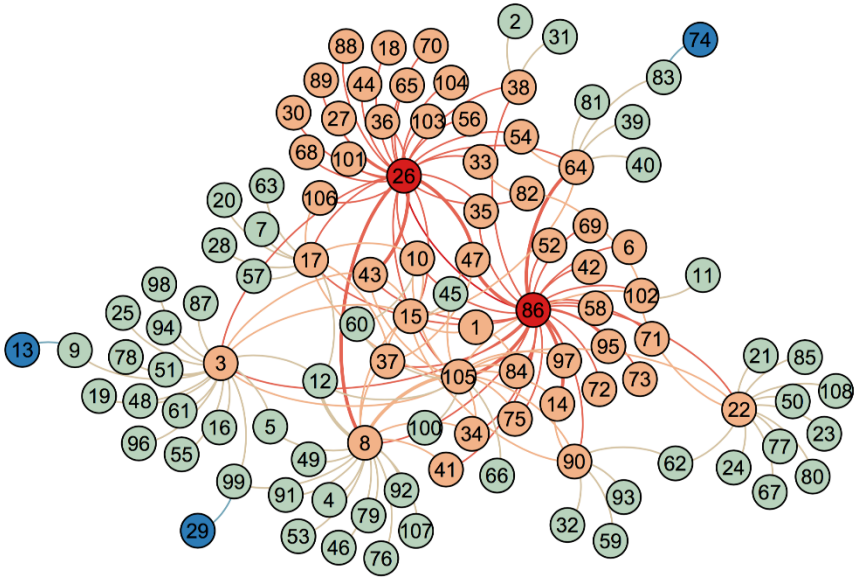
Nodes: Individuals or Groups

Edges: Relationships, interaction



Measurements





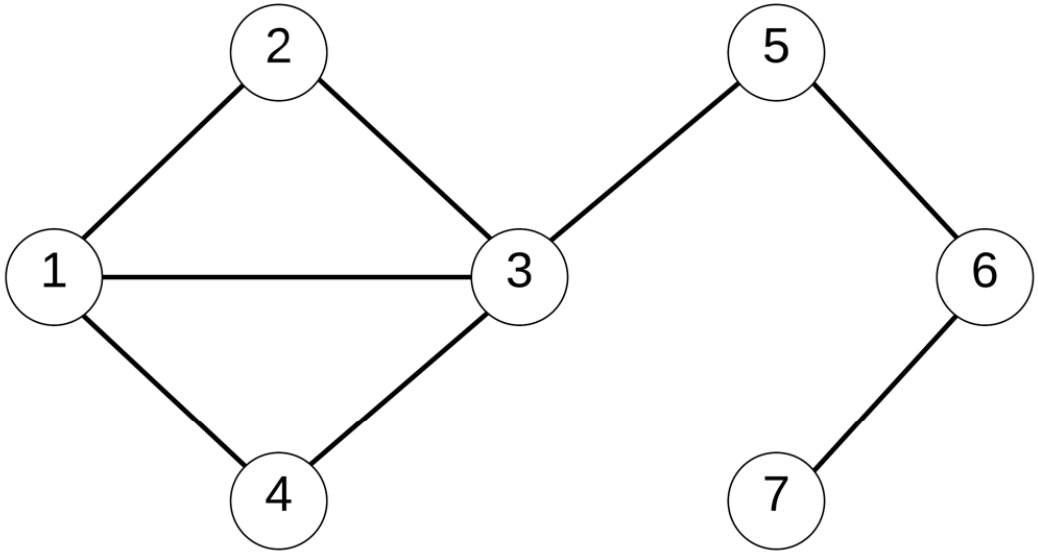
Network-centric measurements

Network size

Counting number of edges/nodes

+ simple

- Not significant, just describing the dimension

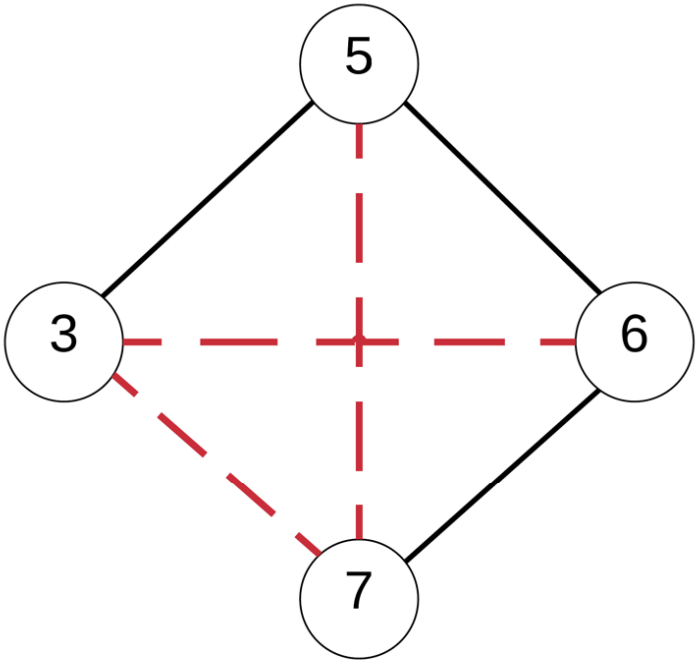


Networksize(nodes): 7

Networksize(edges): 8

Network compactness

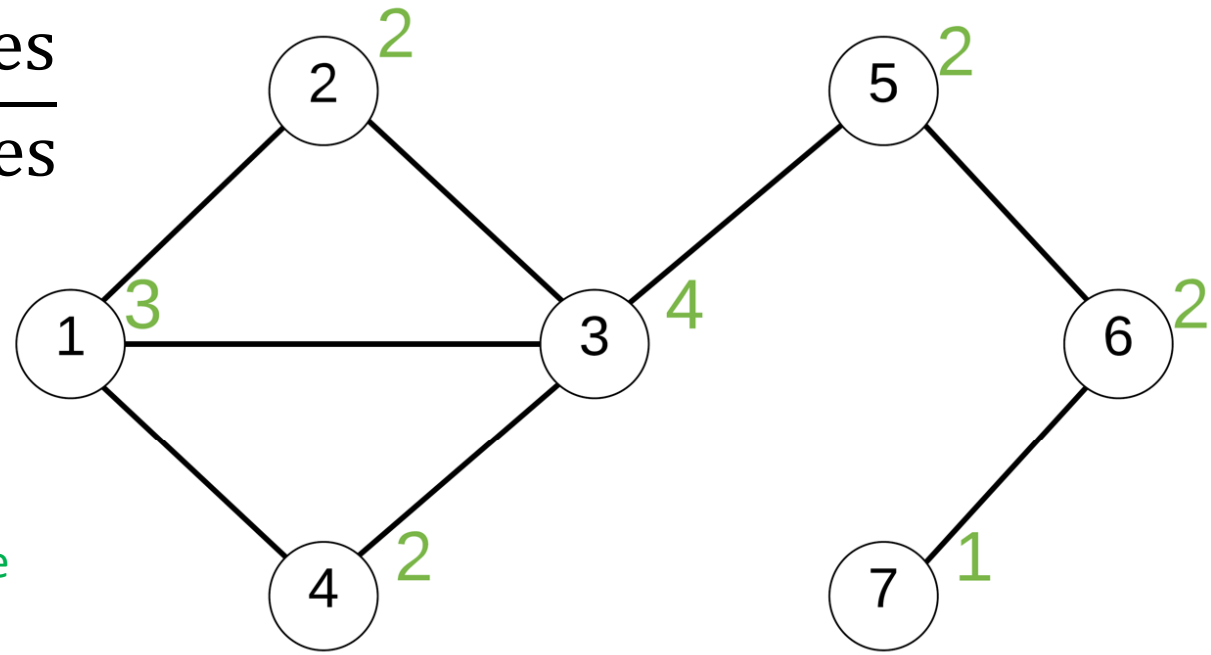
Edges/possibly existing edges
+ describes comparative compactness
+ratio → possible to compare
- Only a general view, no statement about specific node/ area



3 edges (black)
<u>6 possibly existing edges (black & red)</u>
Network compactness: 3/6 = 0.5

Average degree

$$\frac{\text{sum of all degres}}{\text{number of nodes}}$$



+ describes comparative compactness/cohesion
+Node interconnectivity
compare to average → node centric!

(3+2+2+4+2+2+1)/7
≈2.3

Diameter

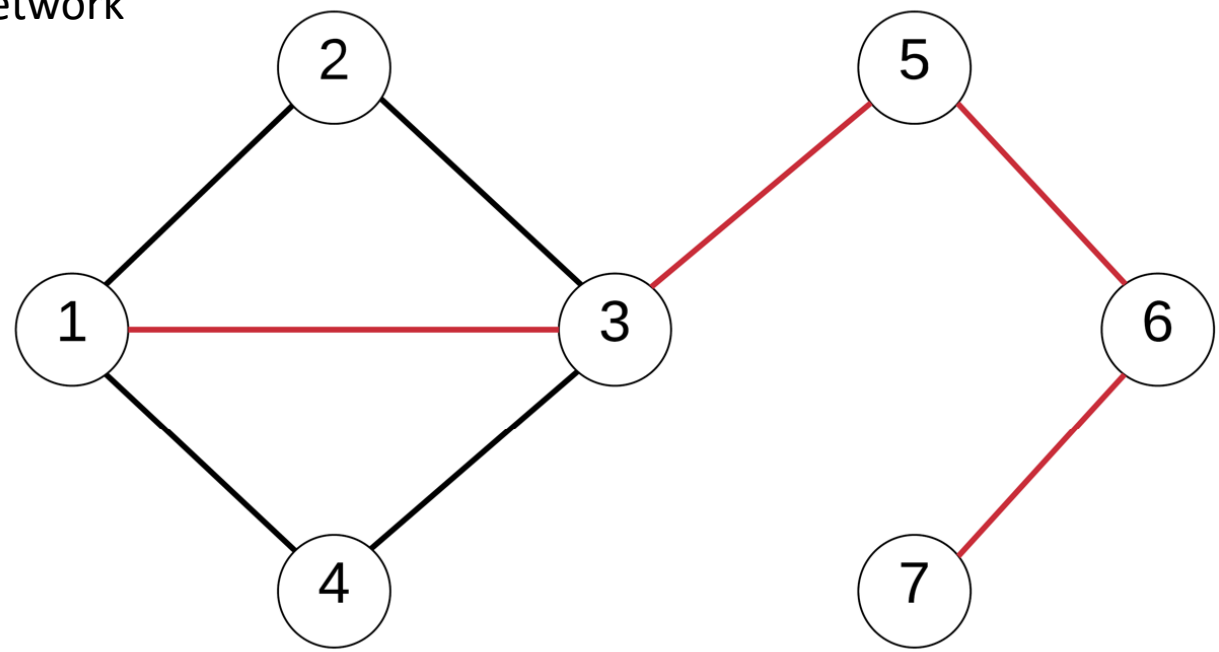
Longest shortest path in network

$\delta = \max\{s(i,j)\}$

S(i,j): shortest path between any two nodes i,j

+measurement of distance in network

-scales with the number of nodes → no comparision possible for different sized networks

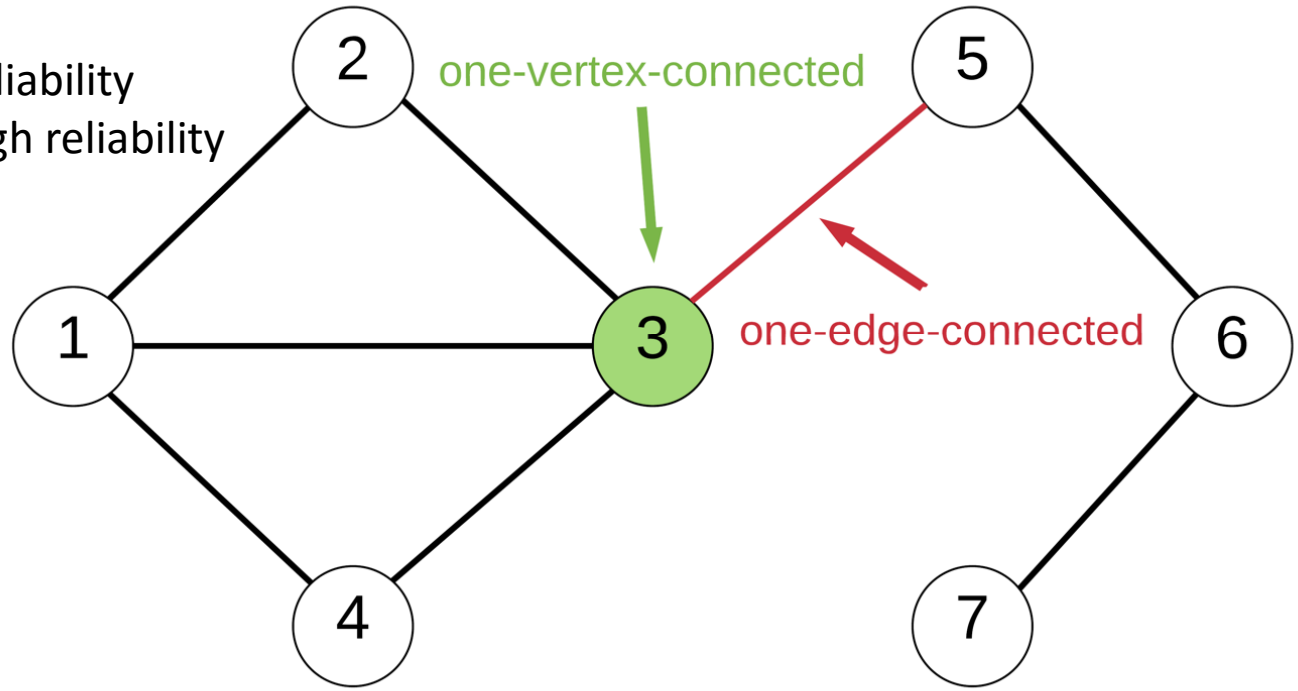


Diameter: 4

Measures of Connectivity

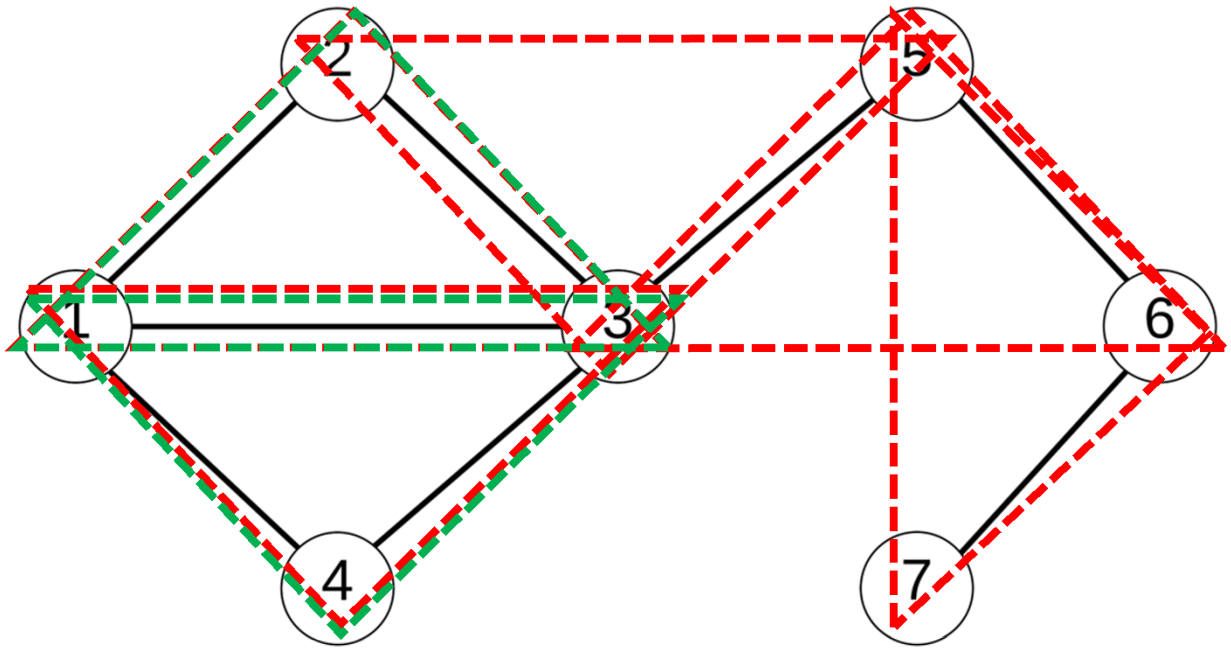
How many edges or nodes to remove until it falls in multiple parts?
→ Searching weakest Link

Describes cohesion/reliability
→high number → high reliability



Global Clustering coefficient

Describes ratio between **triangles** and **triplets**
Range: [0,1]



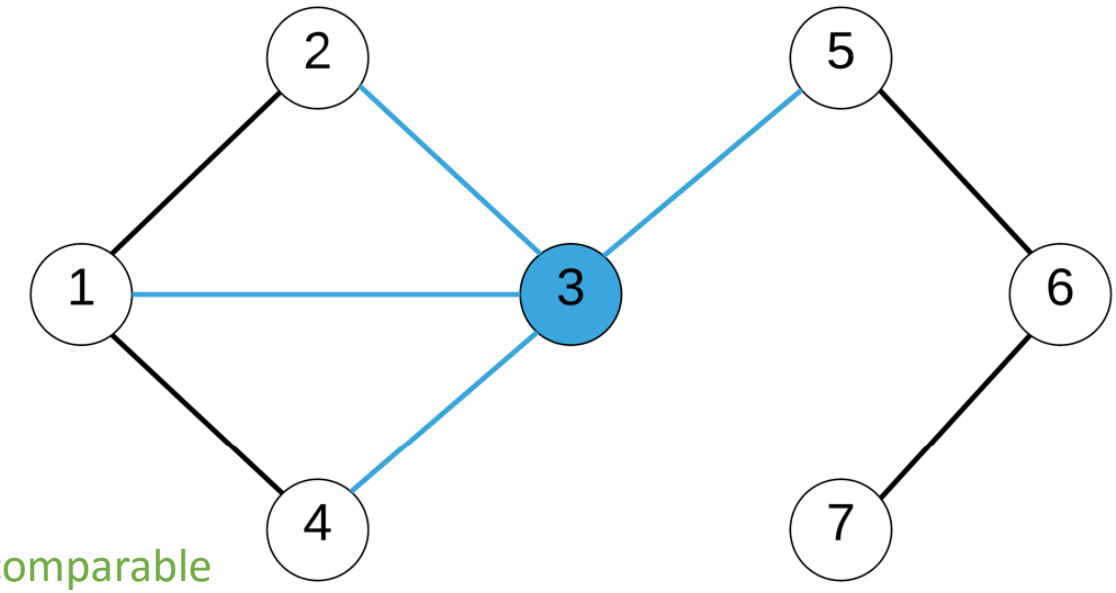
High global clustering
→ good connectivity

2/5 = 0.4

Node-centric measurements

Degree

Number of edges connected to node



+simple
+number of connectivities → comparable

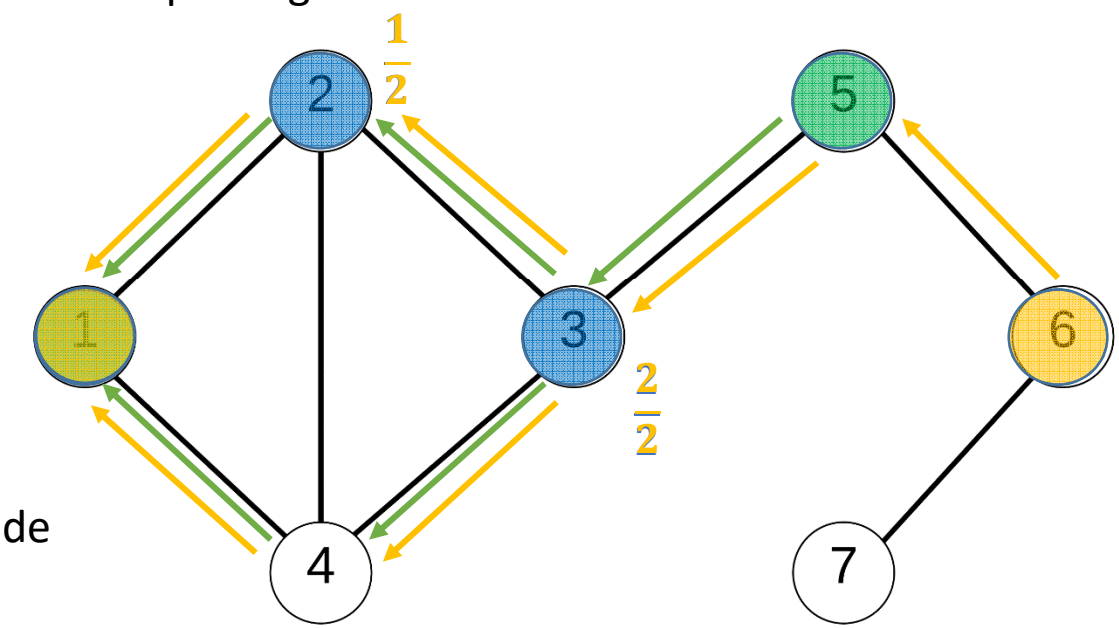
- No information about importance of the connectivities

Node 3: Degree of 4

Betweenness centrality

Sum of all shortest paths connecting two nodes, passing the measured node, divided by all shortest path connecting the same two nodes, including the shortest paths not passing the measured node

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



High number → important node

+ importance of one node

-Scales with number of nodes

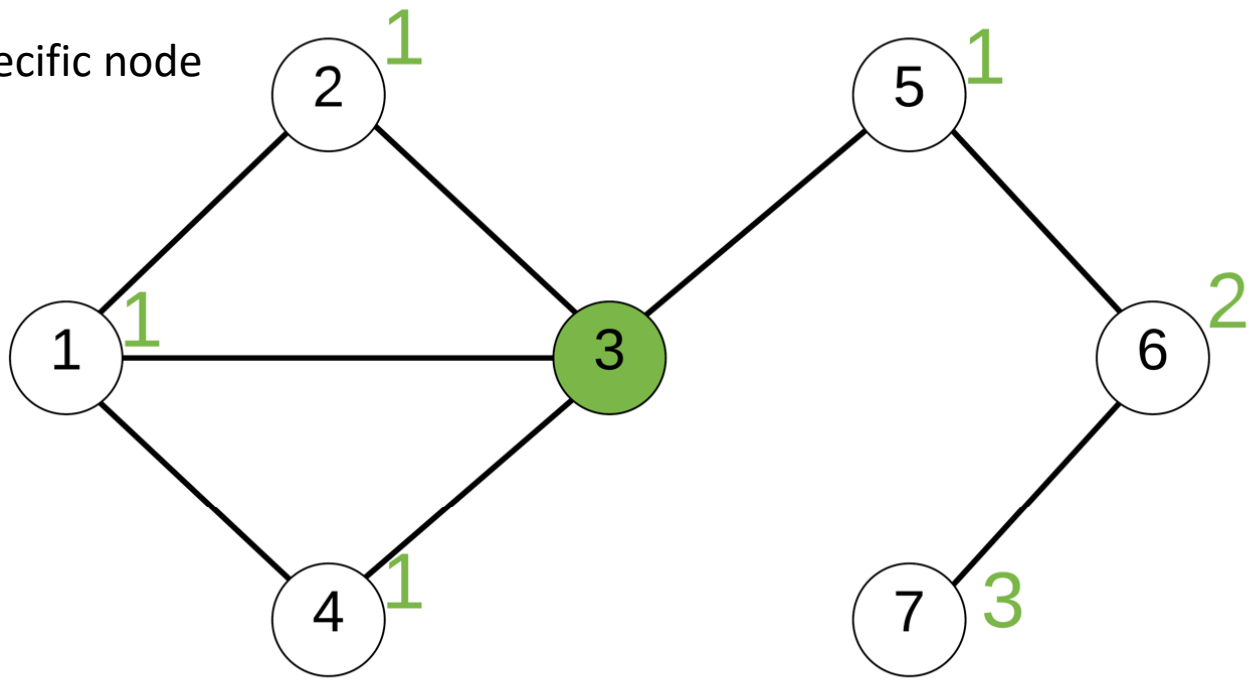
→ No comparision between different networks → Divide by number of nodes

Node 2: $\frac{1}{2} + \frac{1}{2}$ Node 3: $\frac{2}{2} + \frac{2}{2}$

Closeness centrality

Inverse of farness from one node to all other

Showing centrality of a specific node



$$C(x) = \frac{1}{\sum_y d(y,x)}.$$

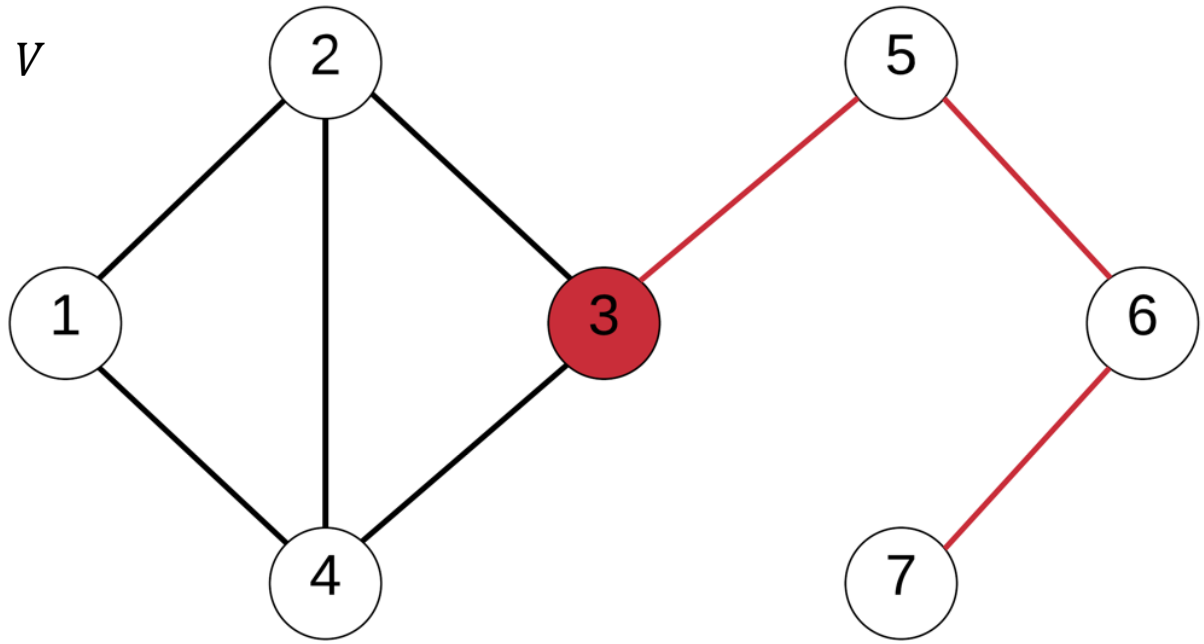
Closeness centrality:
1/(1+1+1+1+2+3) ≈ 0.11

Eccentricity

Number of longest shortest path for a node

$$e(u) = \max d(y, x) : x \in V$$

e(u): eccentricity
d(y,x): shortest path y→x



→ How far from the furtest other?

-scales with number of nodes

Eigenvector centrality

- Assings relative score to a node
- High scoring neighbours → raise the score of the node

- Measurement of influence

Examples: Google PageRank, Katz centrality

Local Clustering coefficent

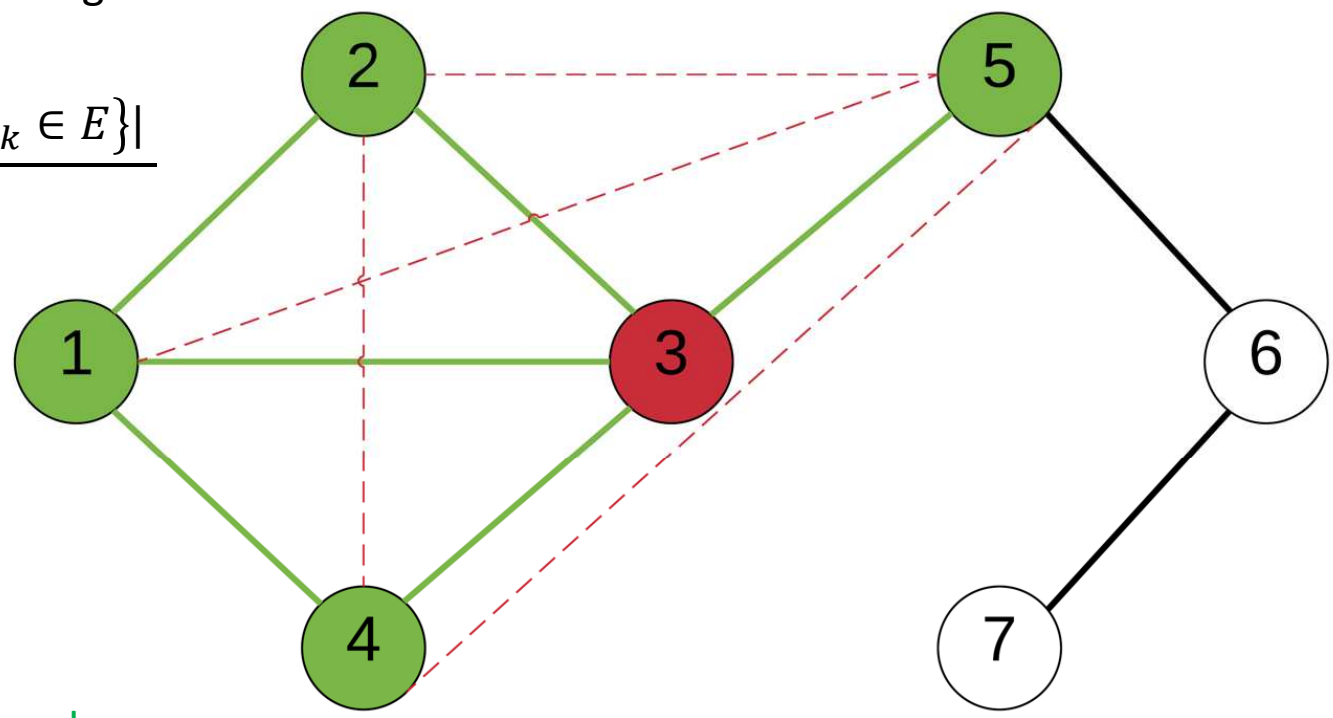
$$\frac{\text{Edges in neighborhood}}{\text{possibly existing edges in neighborhood}}$$

$$C_i = \frac{2|\{e_{jk}:\ v_{j,v_k} \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}$$

Num of max edges in N:

$$\frac{(k_i - 1)}{2} k_i$$

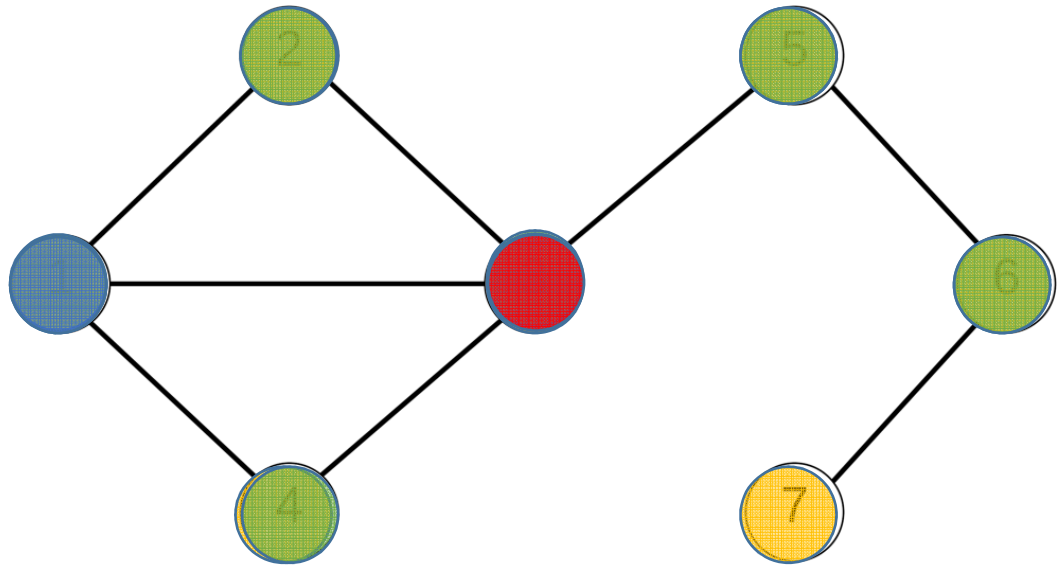
+Comparative number describing clustering of node



Local clustering coefficent:
6/10

Coreness, k-core

Largest subgraph of connected nodes, where each node has degree of at least k



Rank of node: combination of degree and centrality

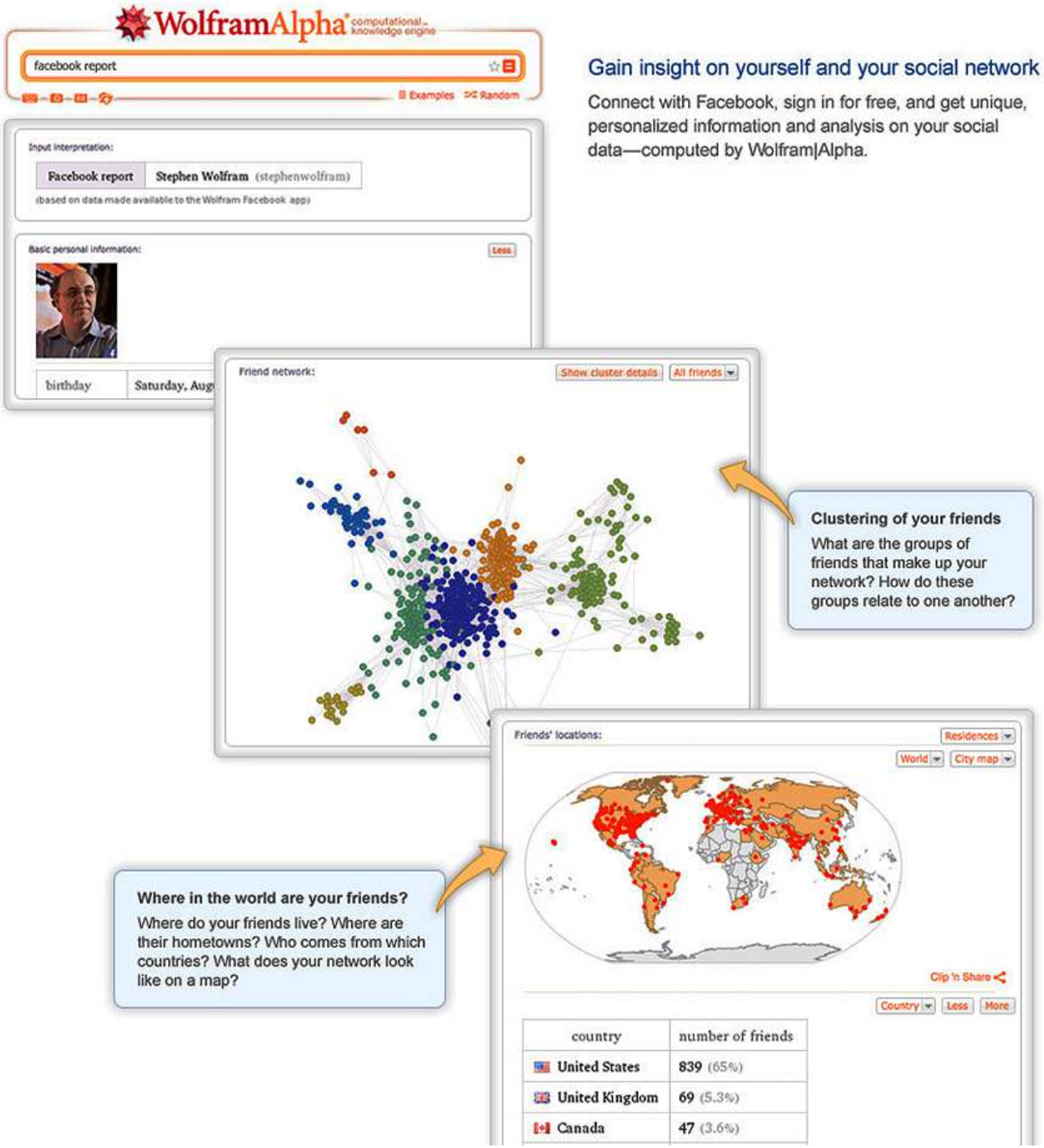
1-core Subgraph
2-core Subgraph
3-core Subgraph
4-core Subgraph

Comparison of Centrality Measurements

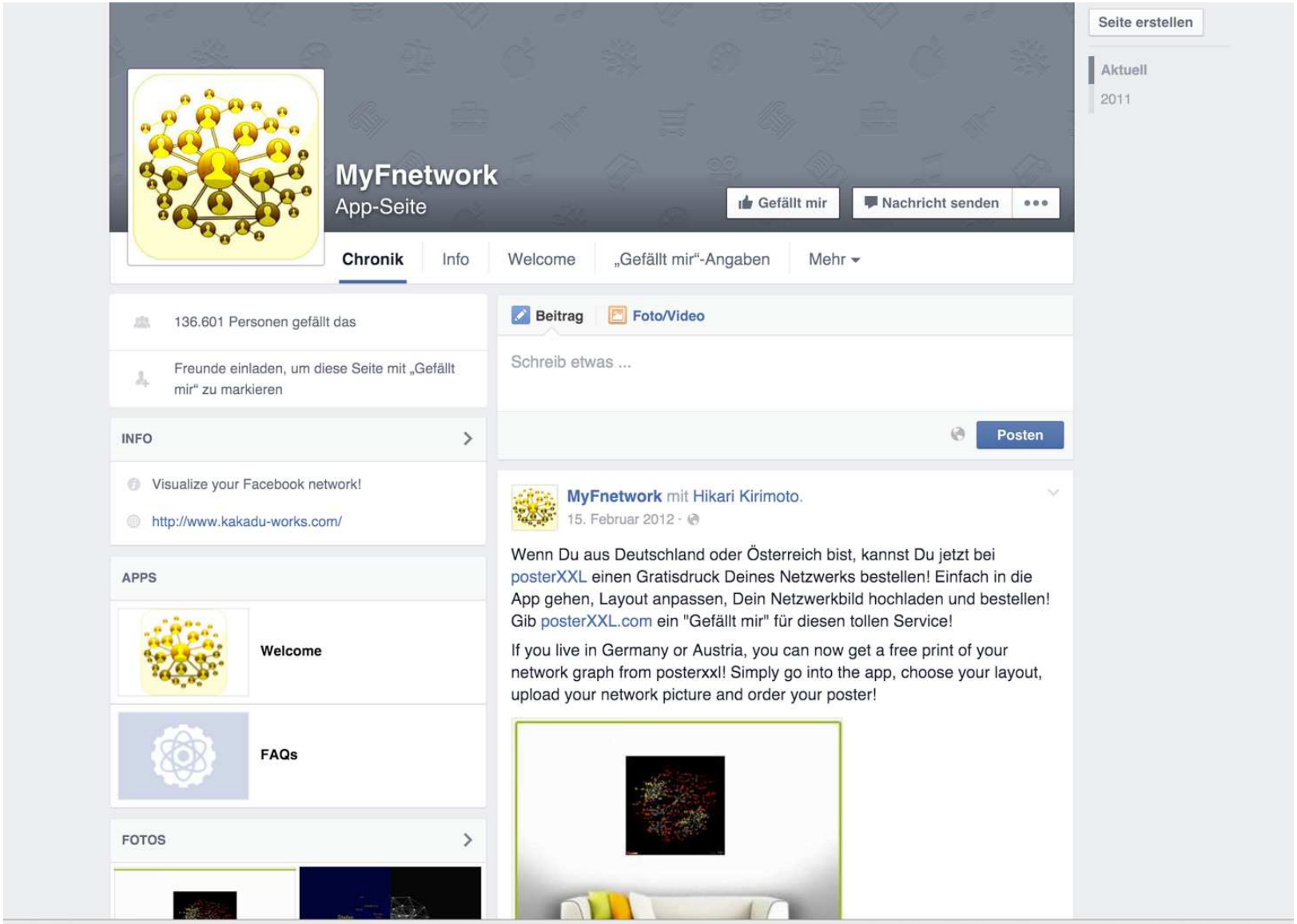
	Low Degree	Low Closeness	Low Betweenness
High Degree		Key player tied to important important/active alters	Ego's connections are redundant - communication bypasses him/her
High Closeness	Key player tied to important important/active alters		Probably multiple paths in the network, ego is near many people, but so are many others
High Betweenness	Ego's few ties are crucial for network flow	Very rare cell. Would mean that ego monopolizes the ties from a small number of people to many others.	

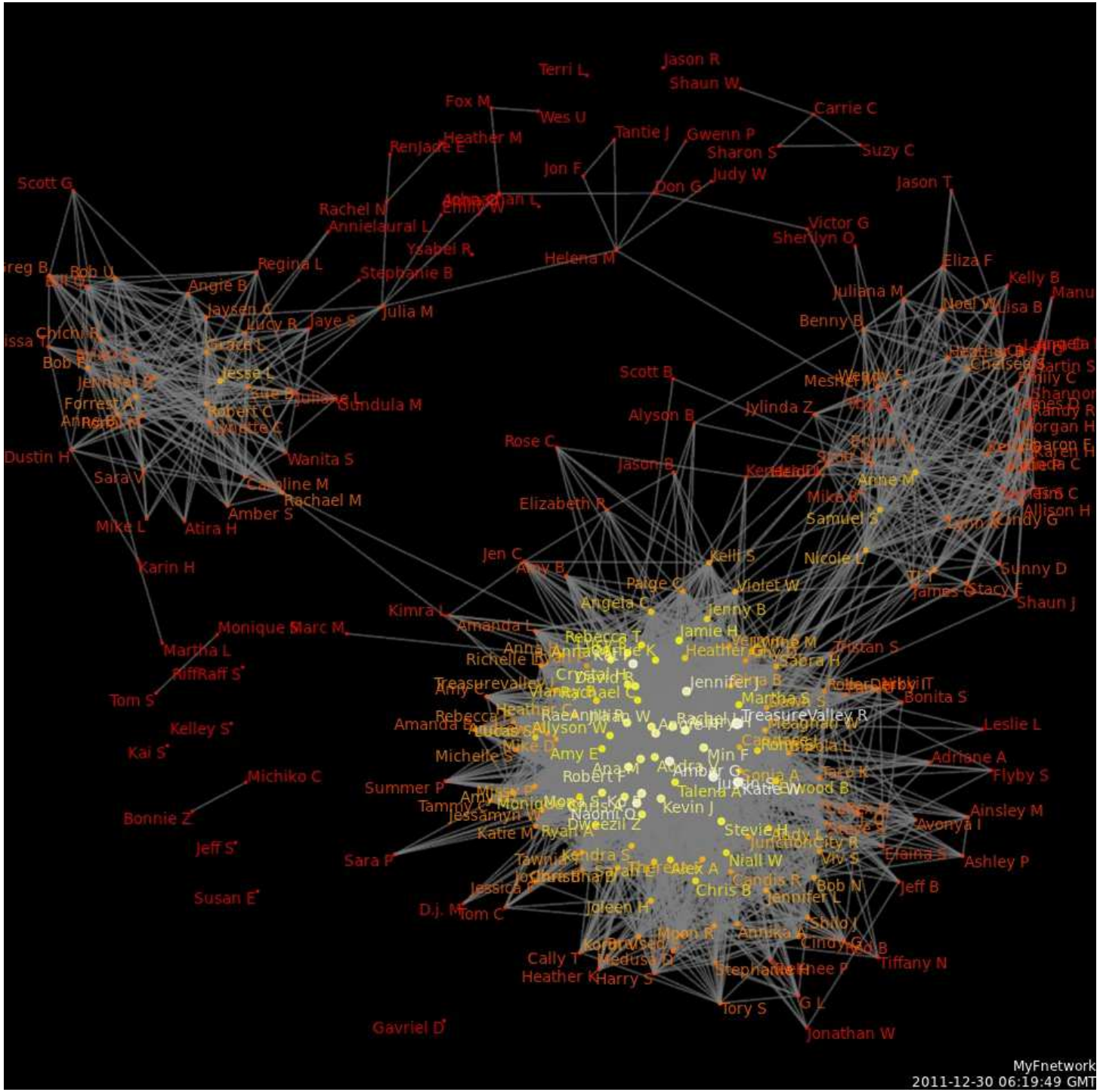
Applications of SNA

SNA

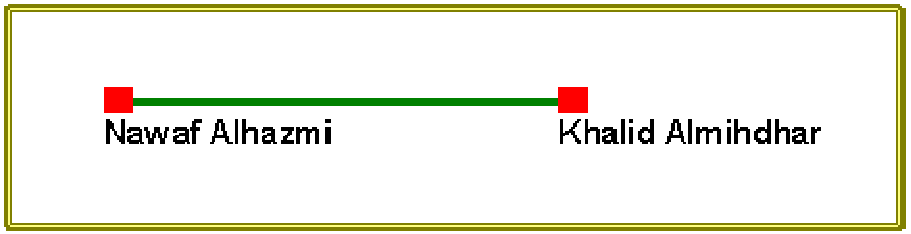

WolframAlpha computational knowledge engine
facebook report
Examples
Random
Input interpretation:
Facebook report
Stephen Wolfram (stephenwolfram)
(based on data made available to the Wolfram Facebook app)
Basic personal information:
Less
birthday
Saturday, Aug
Friend network:
Show cluster details
All friends
Friends' locations:
Residences
World
City map
Clip 'n Share
Country
Less
More
country
number of friends
United States
839 (65%)
United Kingdom
69 (5.3%)
Canada
47 (3.6%)
Gain insight on yourself and your social network
Connect with Facebook, sign in for free, and get unique, personalized information and analysis on your social data—computed by Wolfram|Alpha.
Clustering of your friends
What are the groups of friends that make up your network? How do these groups relate to one another?
Where in the world are your friends?
Where do your friends live? Where are their hometowns? Who comes from which countries? What does your network look like on a map?

Facebook


Seite erstellen
Aktuell
2011
MyFnetwork
App-Seite
Gefällt mir
Nachricht senden
Chronik
Info
Welcome
„Gefällt mir"-Angaben
Mehr
136.601 Personen gefällt das
Freunde einladen, um diese Seite mit „Gefällt mir" zu markieren
Beitrag
Foto/Video
Schreib etwas ...
Posten
INFO
Visualize your Facebook network!
http://www.kakadu-works.com/
APPS
Welcome
FAQs
FOTOS
MyFnetwork mit Hikari Kirimoto.
15. Februar 2012 ·
Wenn Du aus Deutschland oder Österreich bist, kannst Du jetzt bei posterXXL einen Gratisdruck Deines Netzwerks bestellen! Einfach in die App gehen, Layout anpassen, Dein Netzwerkbild hochladen und bestellen! Gib posterXXL.com ein "Gefällt mir" für diesen tollen Service!
If you live in Germany or Austria, you can now get a free print of your network graph from posterxxl! Simply go into the app, choose your layout, upload your network picture and order your poster!

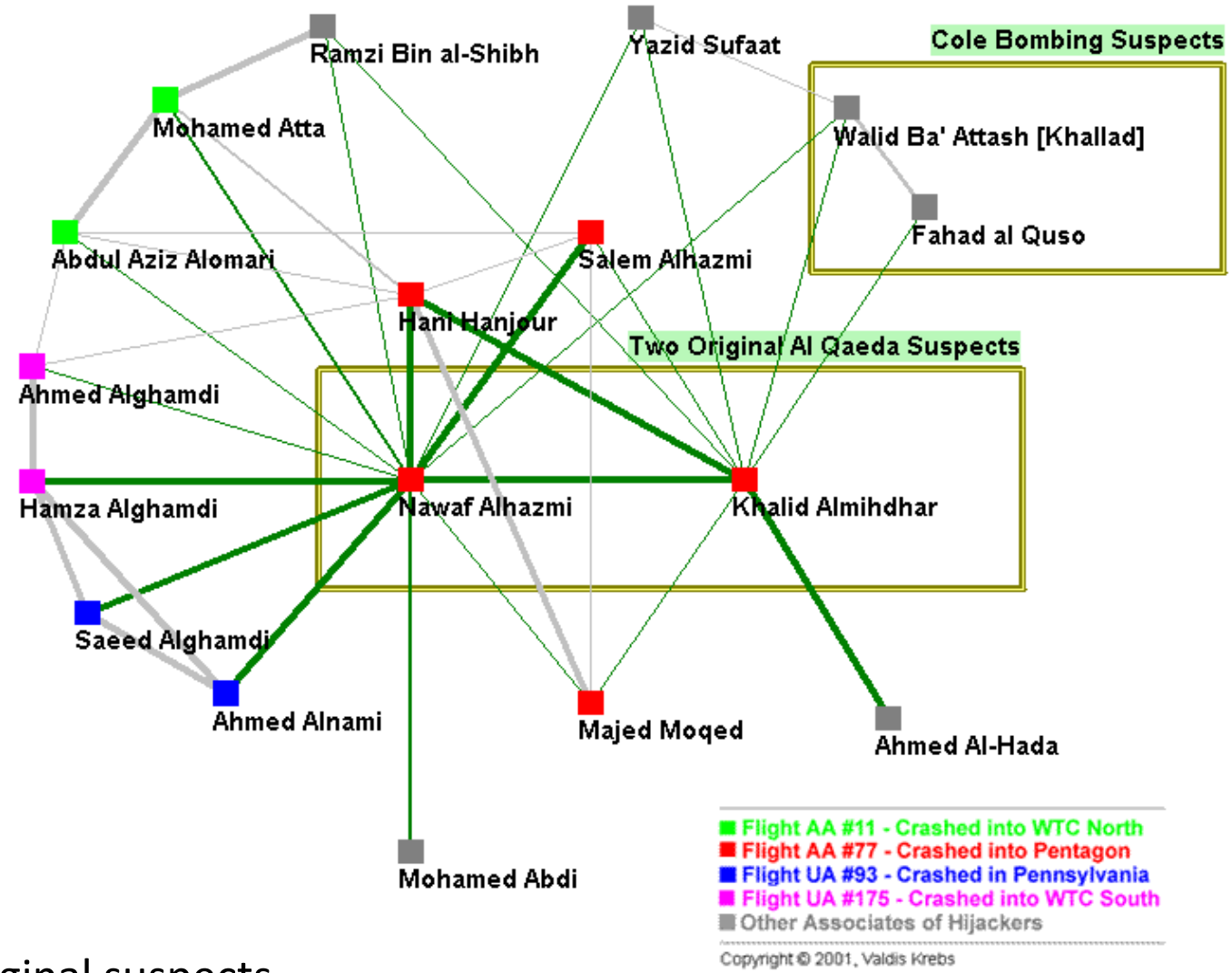

Terri L
Jason R
Shaun W
Fox M
Wes U
Carrie C
Heather M
Tantie J
Gwenn P
Sharon S
Suzy C
Jon F
Judy W
Don G
Jason T
Scott G
Rachel N
Annielaural L
Victor G
Sherilyn O
Ysabel R
Helena M
Eliza F
Regina L
Stephanie B
Kelly B
Angie B
Juliana M
Julia M
Benny B
Jaysen C
Lucy R
Jaye S
Chichi R
Jesse L
Scott B
Gundula M
Jylinda Z
Alyson B
Rose C
Dustin H
Wanita S
Jason B
Sara V
Caroline M
Rachael M
Elizabeth R
Amber S
Mike L
Atira H
Samuel S
Allison H
Jen C
Kelli S
Nicole L
Karin H
Sunny D
Paige C
Violet W
Angela C
Jenny B
Shaun J
Kimra L
Amanda L
Jamie H
Monique S
Marc M
Martha L
Tristan S
Tom S
Bonita S
Leslie L
Kelley S
Kai S
Michiko C
Adriane A
Flyby S
Summer P
Kevin J
Ainsley M
Bonnie Z
Avonya I
Jeff S
Sara P
Ashley P
Susan E
Jeff B
Chris B
Tom C
Tiffany N
Cally T
Heather K
Harry S
Tory S
Gavriel D
Jonathan W
MyFnetwork
2011-12-30 06:19:49 GMT

Social Network Analysis of Terrorist Networks



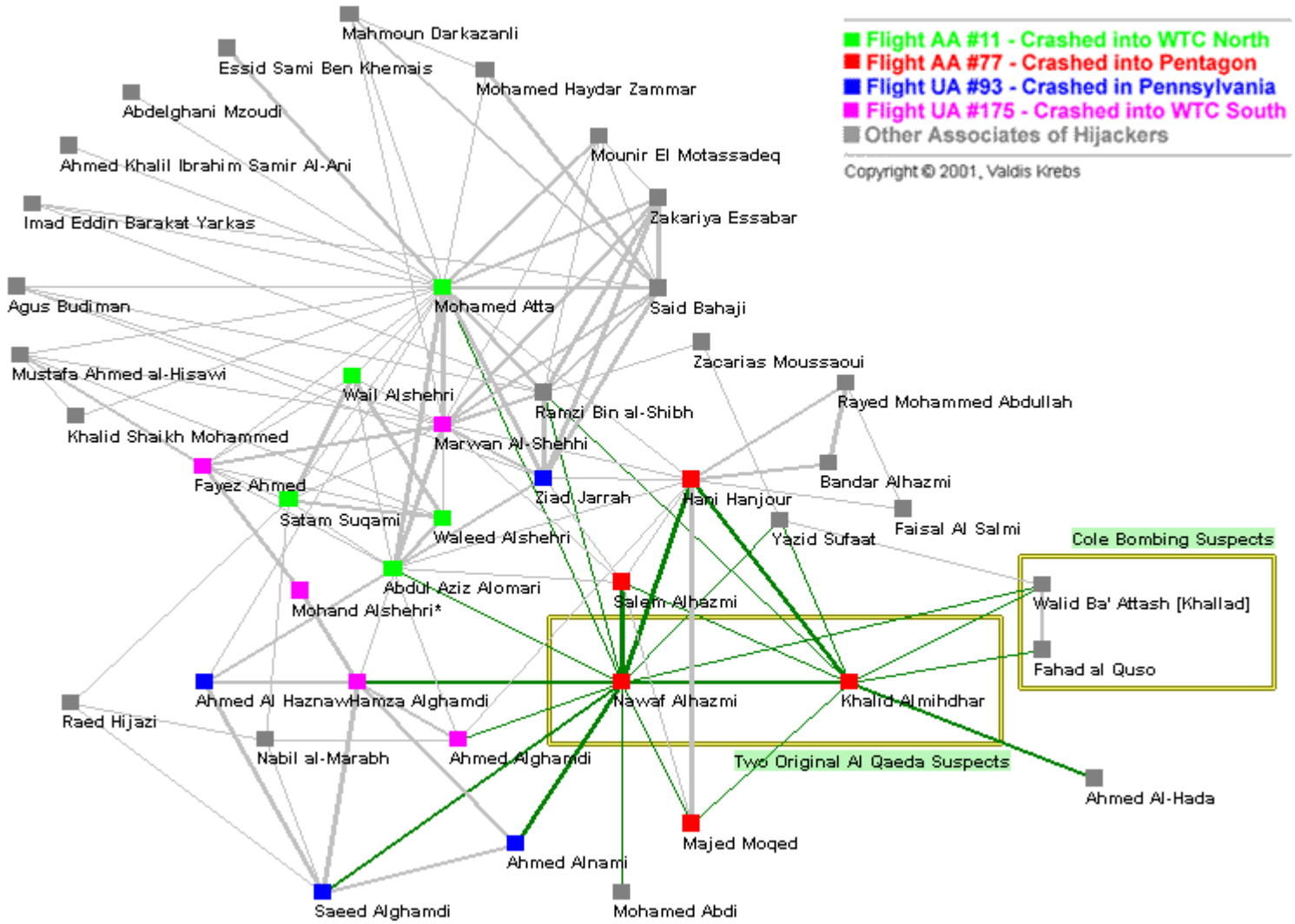
Two initial suspects linked to al-Qaeda

Social Network Analysis of Terrorist Networks



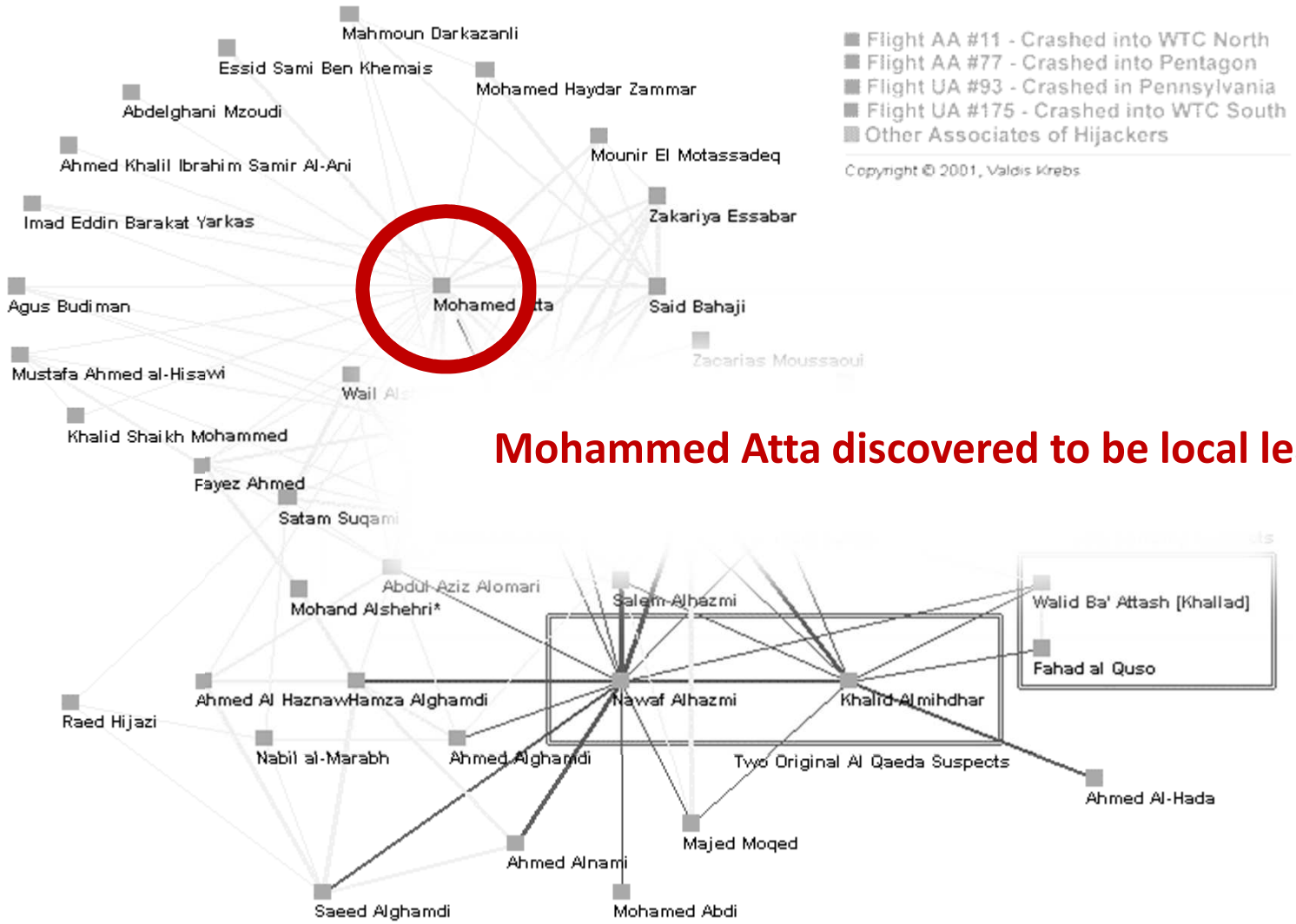
Direct links to original suspects

Social Network Analysis of Terrorist Networks



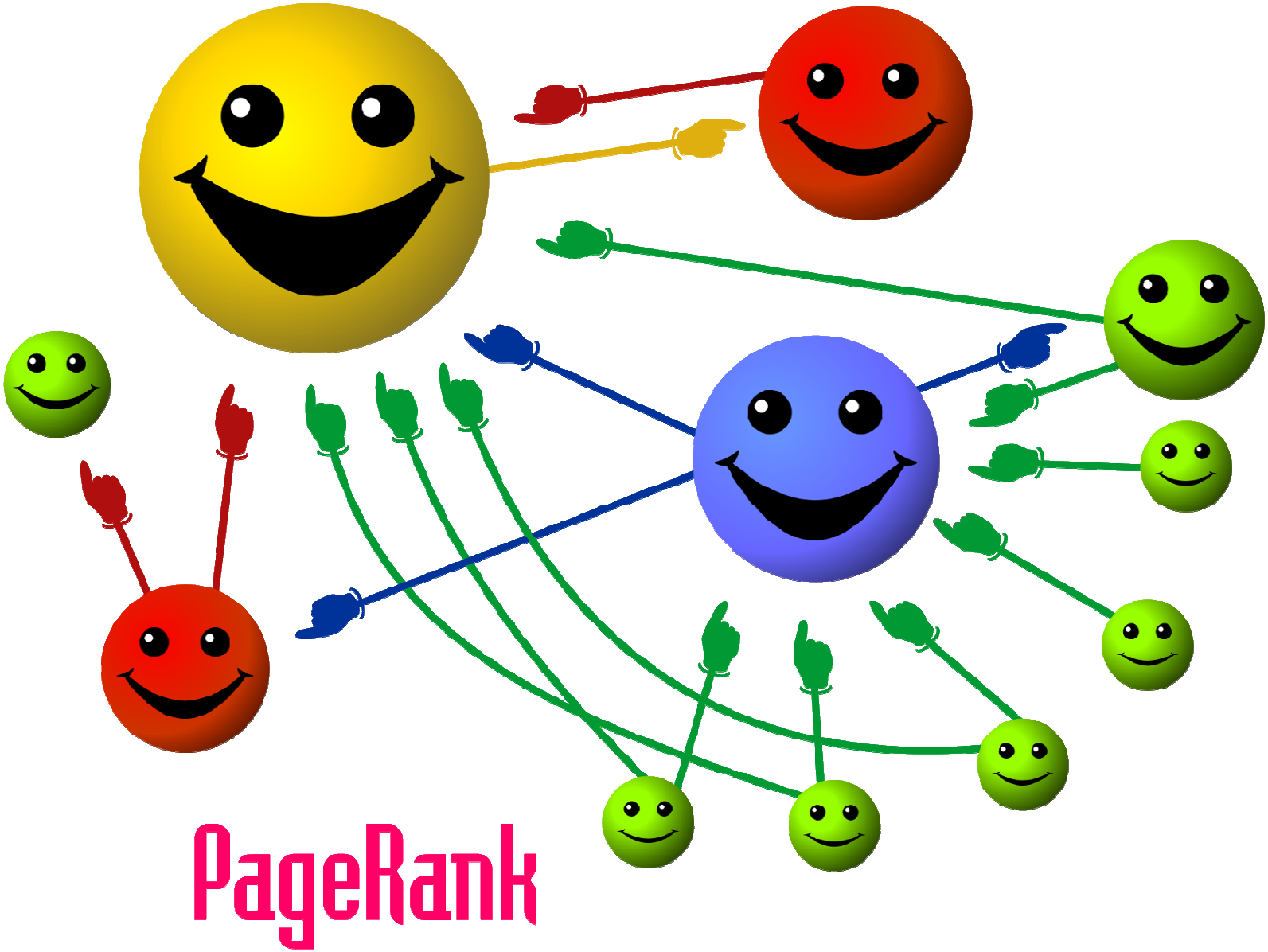
Indirect links to original suspects

Social Network Analysis of Terrorist Networks



Mohammed Atta discovered to be local leader

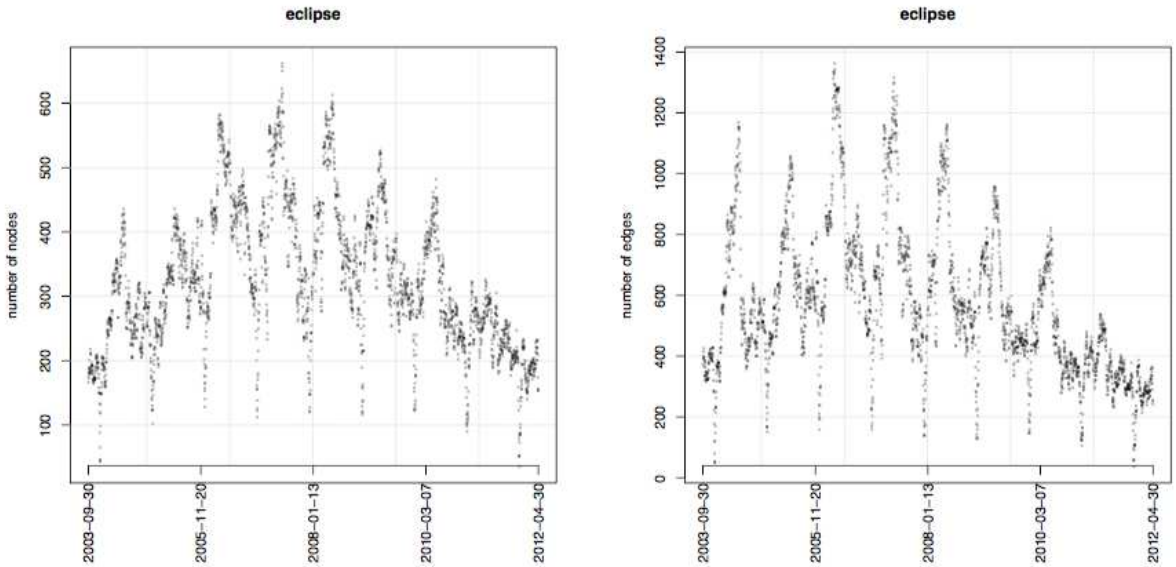
Page Rank



SNA in the Enterprise

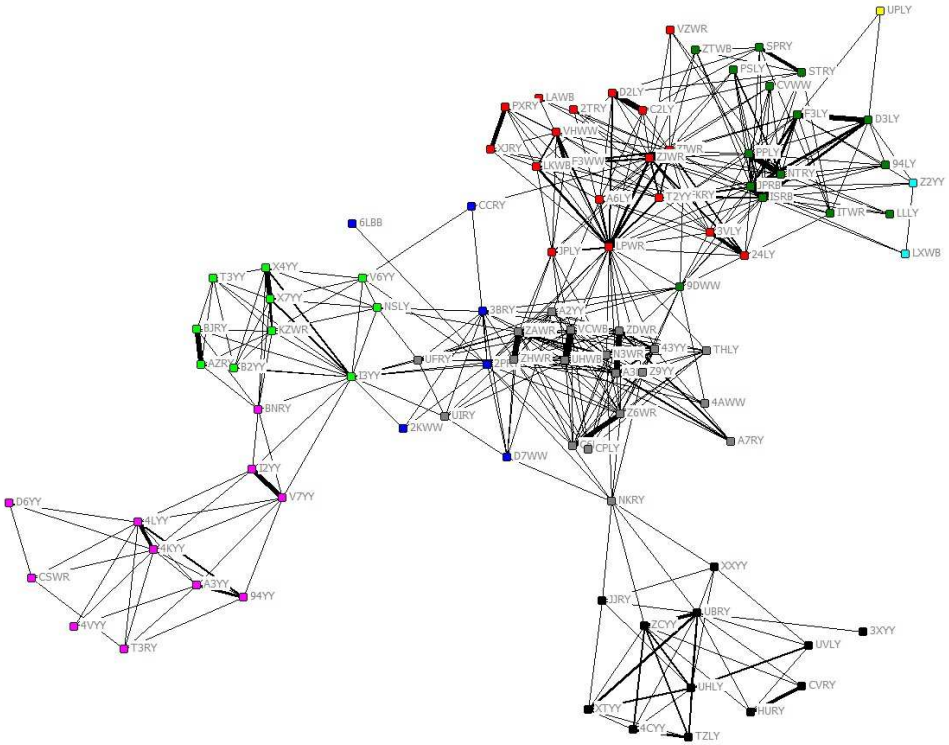
Different possibilities to model a graph

- Time



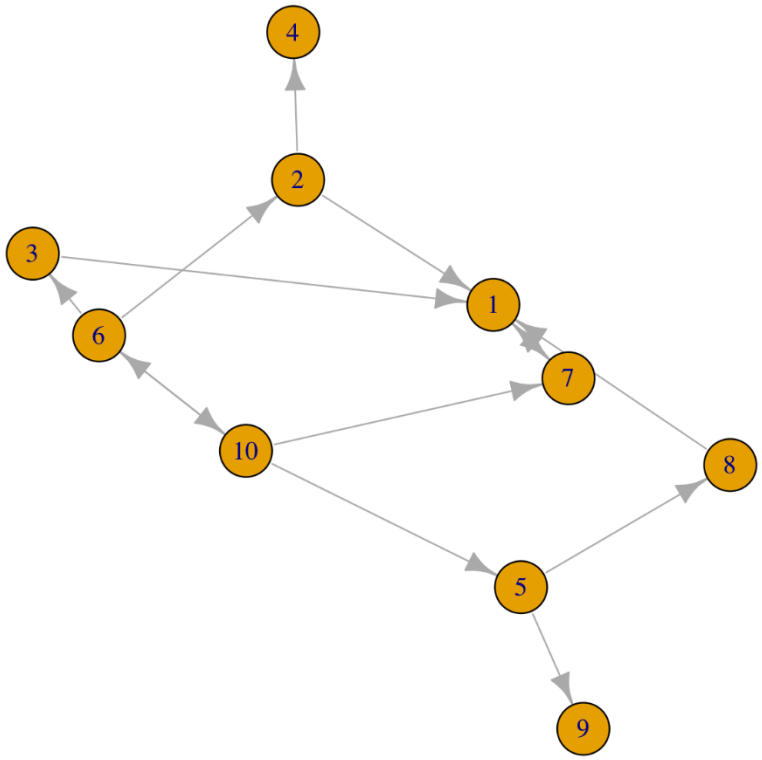
(a) Number of nodes (b) Number of edges

- Weight

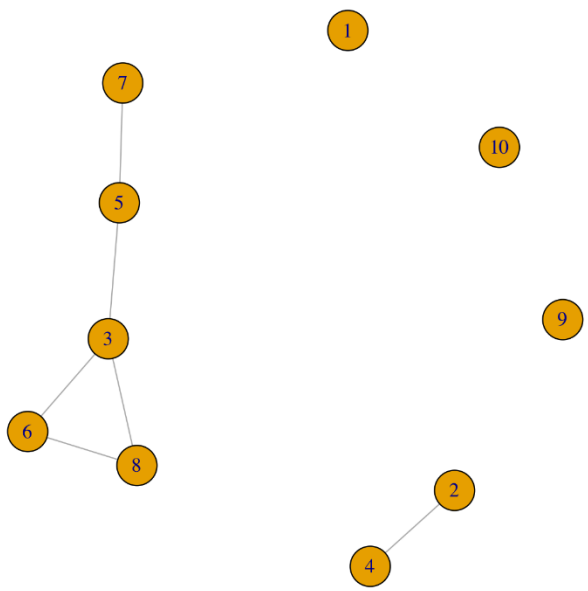


http://irishbrentgoose.blogspot.ch/2012/07/social-networks-revisited.html

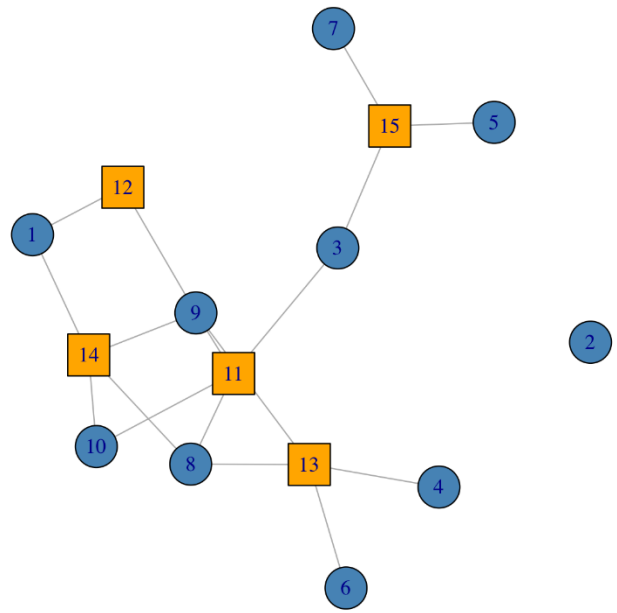
- Directed



One Mode

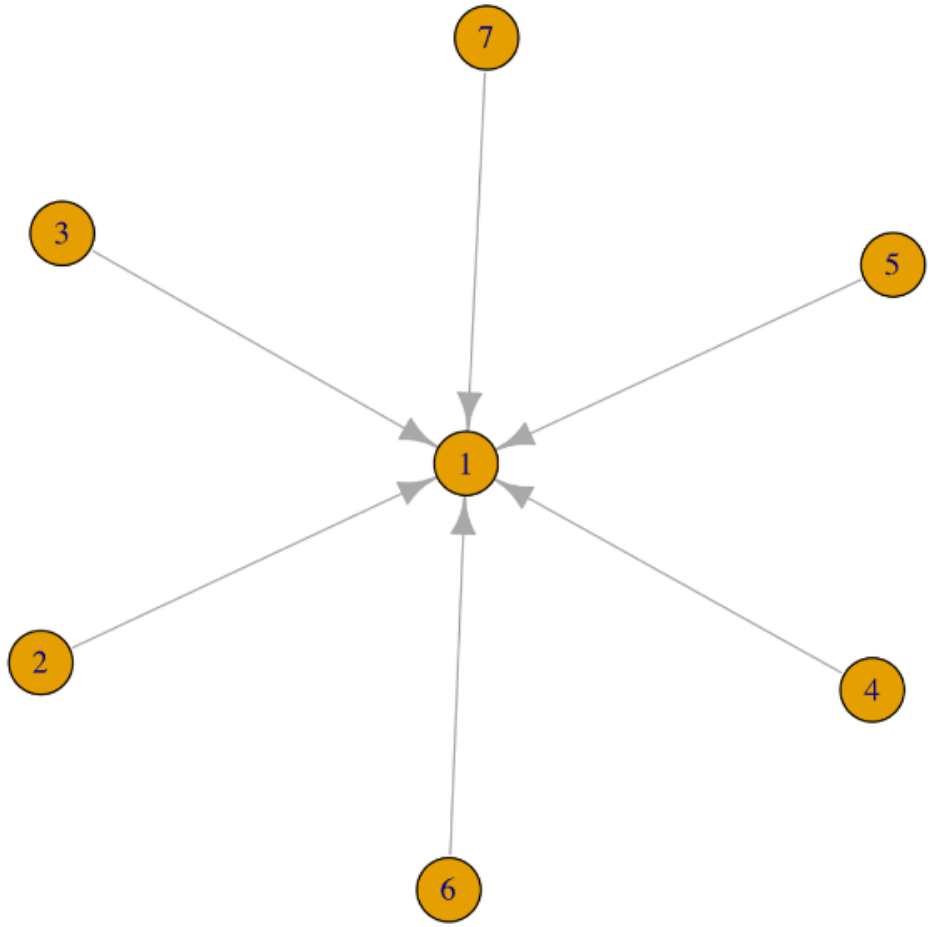


Two Mode

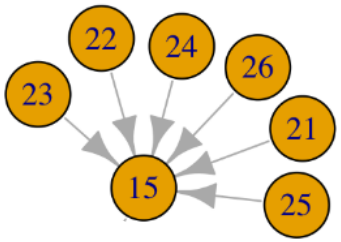


Our Model of the BitTorrent Network

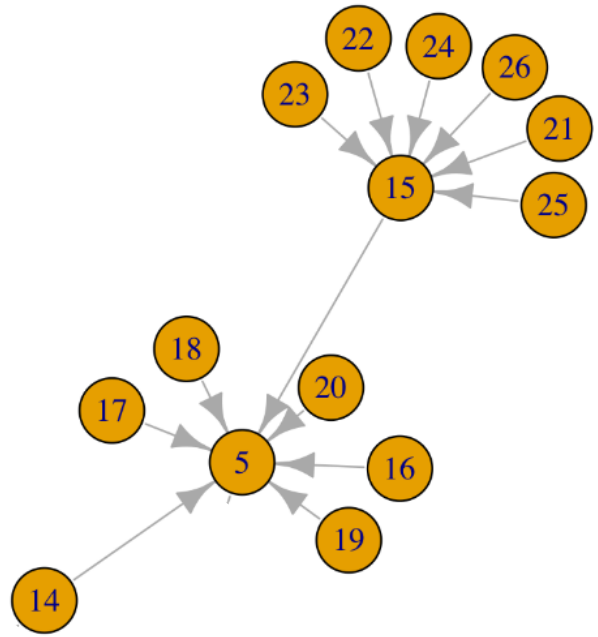
Our model of the BitTorrent Network



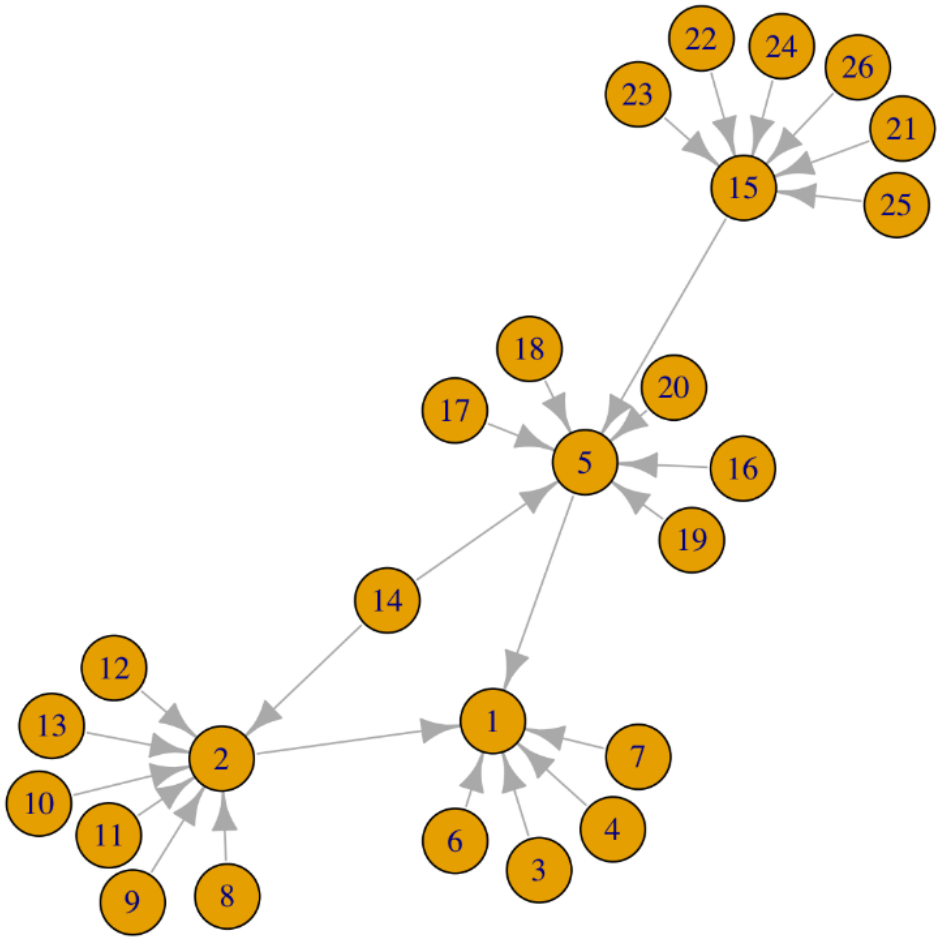
Our model



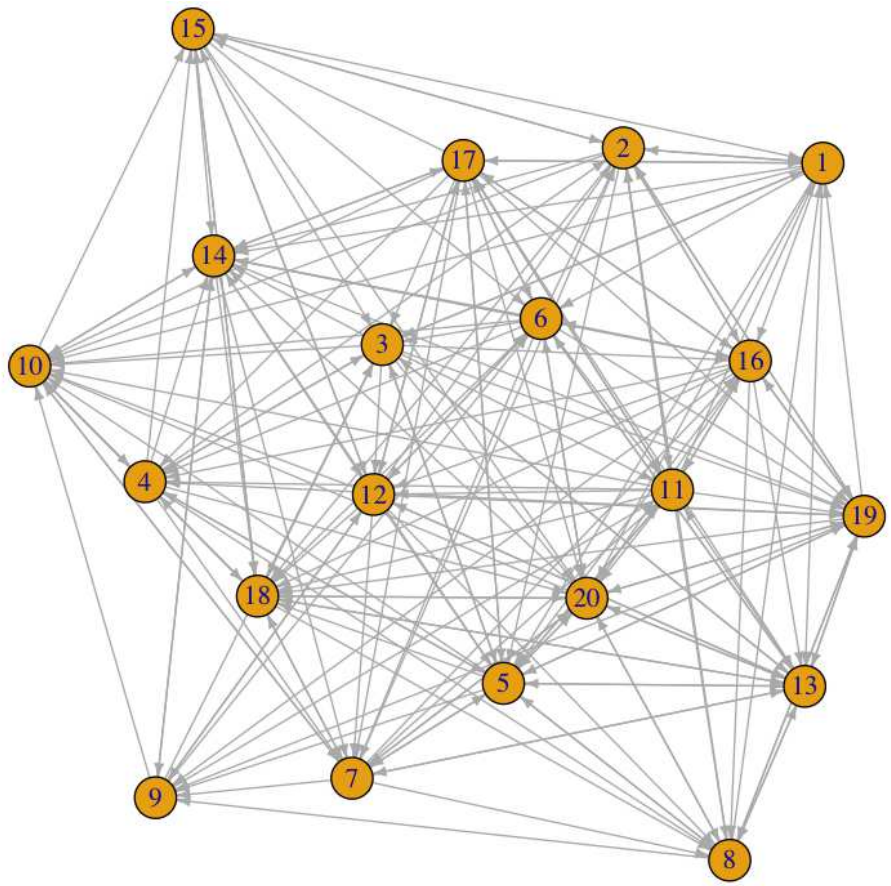
Our model



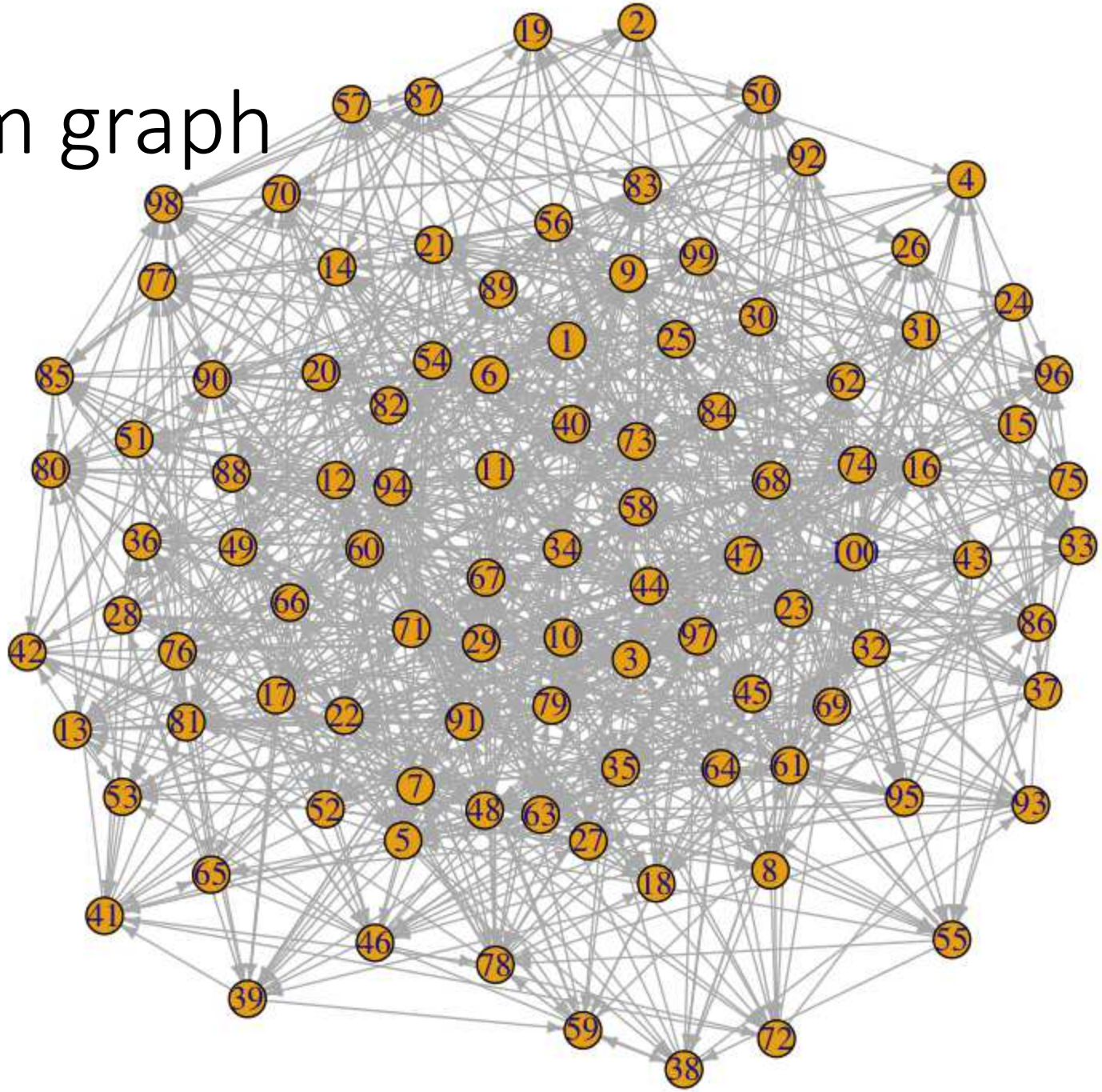
Our model



Random graph according to our model



Random graph



Our model

- Directed
- One mode
- Edge = unspecified number of chunks of a known file
- Nodes = Peers

Possible enhancements

- weighted

Some interpretations of the measurements

Interpretation of the measurements

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Cetrality
- Clustering Coefficient

Interpretation of the measurements

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Cetrality
- Clustering Coefficient

Interpretation of the measurements

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Cetrality
- Clustering Coefficient

Interpretation of the measurements

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Cetrality
- Clustering Coefficient

Interpretation of the measurements

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- Eigenvector Cetrality
- Clustering Coefficient

Optimization

Optimization

- Performance
  - Tracker Localized Algorithm
  - Piecepicker Localized Algorithm
  - Friend list approach

Optimization

- System Integrity

Optimization

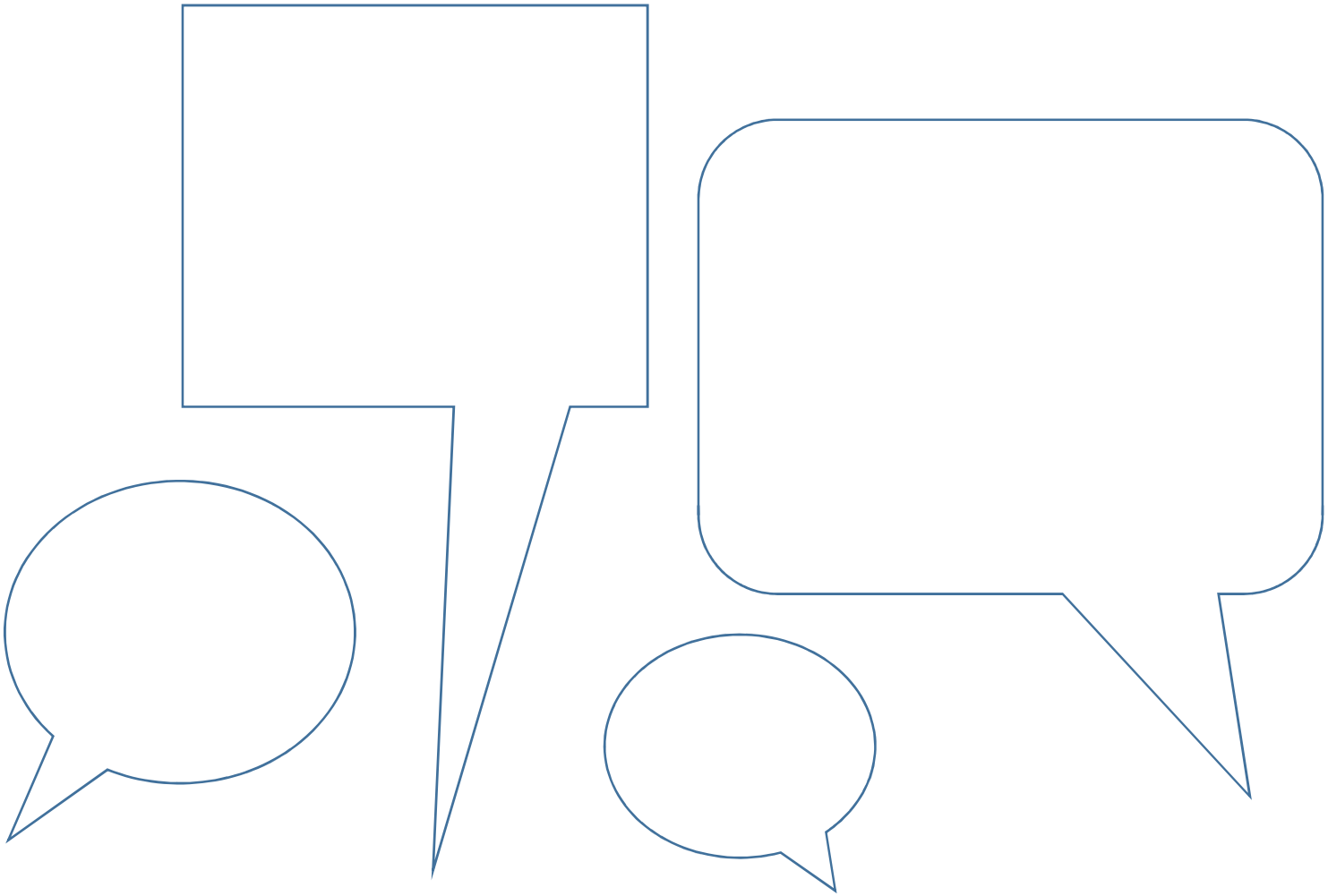
- Free riding

Conclusion

- BitTorrent is not the best P2P system to apply SNA because of the role of the tracker
  - Random Nodes are returned
  - BitTorrent is already a better system then the Beginnings of P2P file sharing systems like Gnutella
- SNA is a very powerful instrument to get insights of structures that are hard to see
- Many measurements depending on the graph model

Questions?

Discussion



Which (if any) P2P systems do you use and why?

Did you experience problems such as free-riding?

What do you think about free riding in BitTorrent?

Is it ok to only consume and not contribute?

Do you see weaknesses in our model how we modeled the graph of the file distribution systems in BitTorrent?

What would you change?

As we heard from Benjamin SNA's might be used to enhance the social network in enterprises

e.g. By adding new edges

Do you see problems with that?

Do you th dds or
diminishe cebook?

What is your opinion about SNA/Information gathering in Facebook, Google+ etc?

How far is it allowed to go?

Friends count is basically a degree measure in facebook, do you see also a use of a closeness or betweenness centrality, why ? Why not?

Since SNA is a network of relations Could you think of other applications for SNA?

Not in the field of social life?