

Social Media Analytics

Network Centrality Metrics

MSBA, Session 3, Jan 28, 2022

Dr. Anitesh Barua

David Bruton Jr. Centennial Chair Professor of Business
Distinguished Fellow, INFORMS Information Systems Society
University of Texas Distinguished Teaching Professor
Associate Director, Center for Research in e-Commerce
McCombs School of Business, University of Texas at Austin

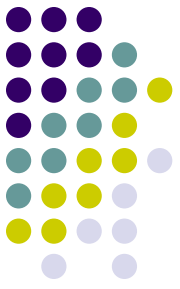
Email: aniteshb@gmail.com





Metric 2: Betweenness Centrality

- Popularity (degree) is useful, but there are other roles
- Connecting *disparate* parts of a network
- How difficult will it be for others (esp. at the “outskirts”) to communicate without you in the network?
- Nodes with high betweenness are important in transmitting new information, ideas & opportunities to a wide audience



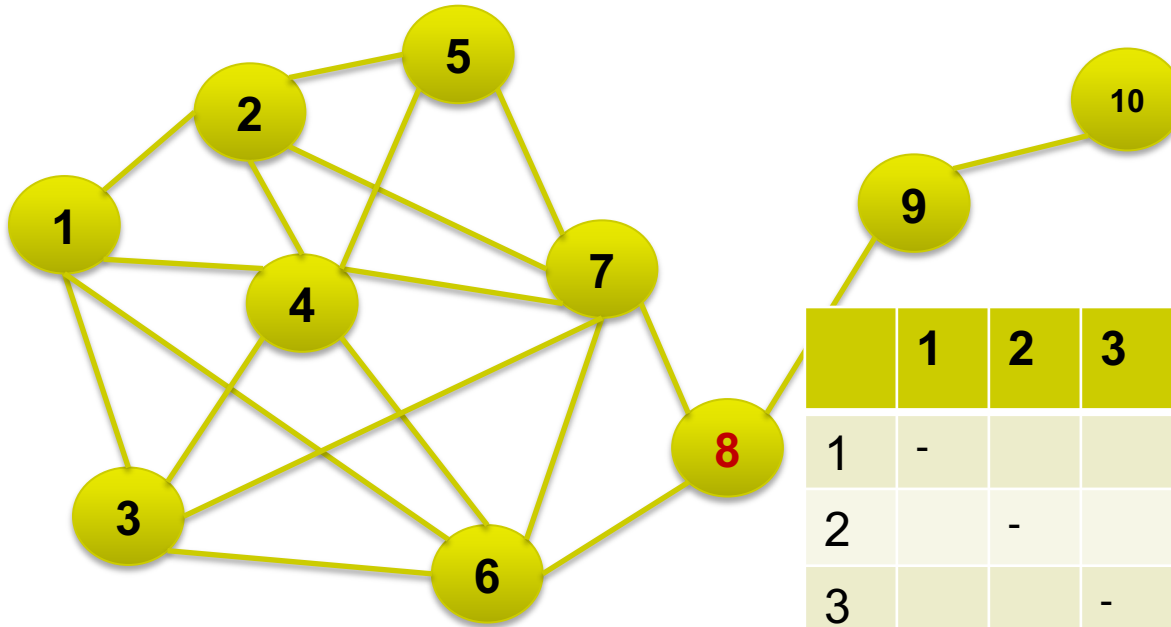
Formal Definition

Let g_{st} be the number of geodesics (shortest paths) between nodes s and t

Let $g_{st}(i)$ be the number of geodesics between s and t passing through node i

Betweenness centrality of node i : $b_i = \sum_{s, t (s \neq i \neq t)} \frac{g_{st}(i)}{g_{st}}$

Betweenness Centrality Example



	1	2	3	4	5	6	7	9	10
1	-								
2		-							
3			-						
4				-					
5					-				
6						-			
7							-		
9								-	
10									-



Network Level Metric

Network betweenness centrality: $Cb = \frac{\sum_{i=1}^N y_i}{y_{max}}$

where $y_i = b^* - b_i$ and y_{max} = the (theoretical) highest possible value of the numerator (only a benchmark)

- What would the theoretical highest be?
- For which type of a network?

Metric 3: Closeness Centrality



- Some nodes can reach the whole network more quickly than other nodes
- How close a node is to all other nodes
- Create a matrix of shortest distances (geodesic) between nodes
- Sum of shortest distances between a node and all other nodes
- $1/\text{sum of shortest distances}$ is the closeness centrality of the node



Formal Definition

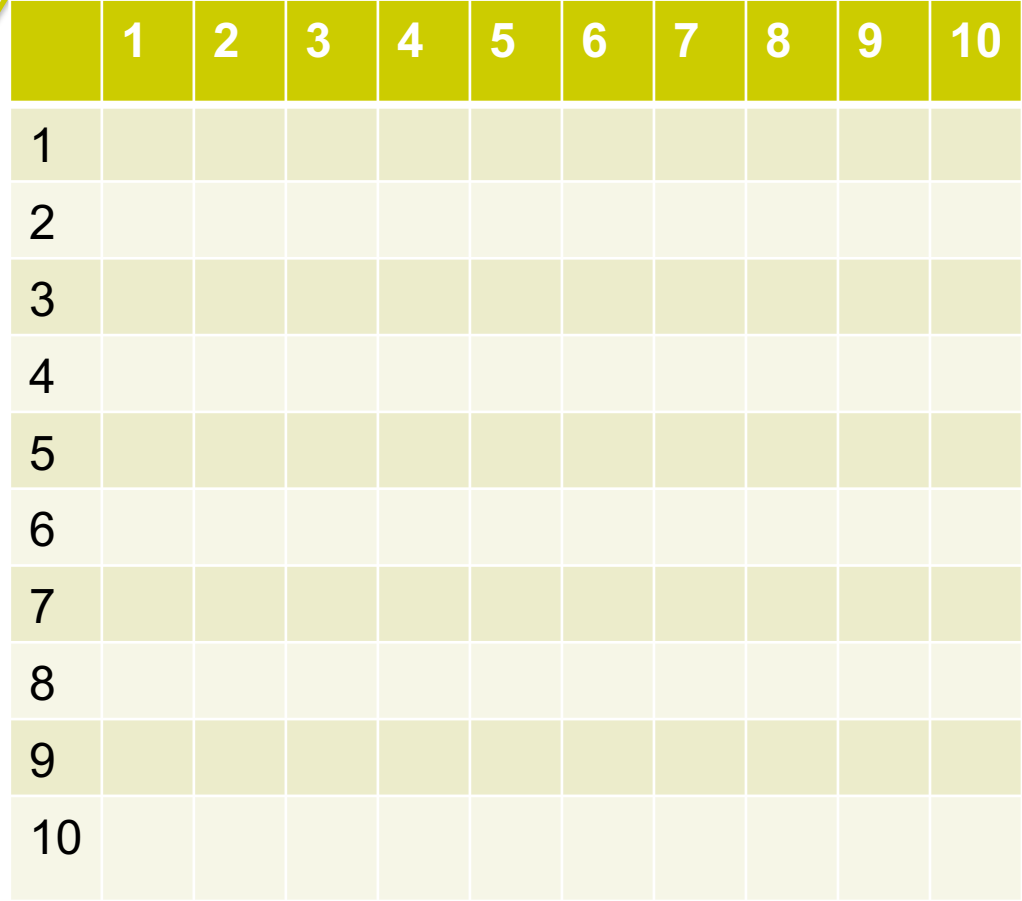
Closeness centrality of node i :
$$c_i = \frac{1}{\sum_{j (j \neq i)} l_{ij}} = \frac{1}{(N-1)\bar{d}_i}$$

Where l_{ij} is the distance between nodes i and j .

\bar{d}_i is the average geodesic length from node i to all other nodes.

What would be the scaled closeness centrality of node i ?

What would be the network level closeness centrality?



The 911 Terrorist Network Revisited

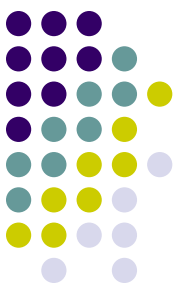
Who are central to the network?
 What was the role of M. Atta?
 How can we watch out against future attacks?



Degrees	Betweenness	Closeness
0.417 Mohamed Atta	0.334 Nawaf Alhazmi	0.571 Mohamed Atta
0.389 Marwan Al-Shehhi	0.318 Mohamed Atta	0.537 Nawaf Alhazmi
0.278 Hani Hanjour	0.227 Hani Hanjour	0.507 Hani Hanjour
0.278 Nawaf Alhazmi	0.158 Marwan Al-Shehhi	0.500 Marwan Al-Shehhi
0.278 Ziad Jarrah	0.116 Saeed Alghamdi*	0.480 Ziad Jarrah
0.222 Ramzi Bin al-Shibh	0.081 Hamza Alghamdi	0.429 Mustafa al-Hisawi
0.194 Said Bahaji	0.080 Waleed Alshehri	0.429 Salem Alhazmi*
0.167 Hamza Alghamdi	0.076 Ziad Jarrah	0.424 Lotfi Raissi
0.167 Saeed Alghamdi*	0.064 Mustafa al-Hisawi	0.424 Saeed Alghamdi*
0.139 Lotfi Raissi	0.049 Abdul Aziz Al-Omari*	0.419 Abdul Aziz Al-Omari*
0.128 MEAN	0.046 MEAN	0.393 MEAN



	Car model	Degree	Betweenness	Eigenvector	No. of occurrences
1	Honda Accord	99	0.954	15.93	58,443
2	Honda Civic	97	0.891	15.77	21,467
3	Lexus ES	97	0.861	15.78	11,540
4	Toyota Camry	97	0.710	15.89	34,559
5	Volkswagen Passat	97	0.794	15.86	16,474
6	Infiniti G35	97	0.713	15.79	23,247
7	Nissan Altima	96	0.605	15.80	13,061
8	Acura TL	95	0.545	15.75	29,400
9	Cadillac CTS	94	0.597	15.56	8,220
10	Nissan Maxima	94	0.533	15.69	10,727
11	Audi A4	94	0.547	15.54	13,454
12	Volkswagen Jetta	94	0.607	15.48	12,251
13	Toyota Corolla	93	0.579	15.41	7,133
14	Chevrolet Impala	92	0.457	15.46	11,659
15	Lincoln LS	92	0.594	15.36	3,092
16	Chrysler 300C	91	0.438	15.35	4,833
17	Toyota Avalon	91	0.597	15.25	12,796
18	Pontiac Grand Prix	91	0.425	15.24	2,327
19	Chevrolet Malibu	90	0.454	15.09	7,235
20	Hyundai Sonata	90	0.397	15.22	16,733
21	Audi A6	89	0.485	14.96	13,617
22	Ford Taurus	88	0.381	15.00	6,907
23	Chevrolet Corvette	88	0.364	14.90	2,254
24	Cadillac STS	87	0.418	14.67	4,071
25	Acura RL	85	0.312	14.65	9,258
26	Chrysler Cruiser	85	0.388	14.48	818
27	Nissan Sentra	85	0.379	14.34	3,170
28	Chrysler 300M	84	0.303	14.43	6,248
29	BMW 3-Series	83	0.355	14.16	2,890
30	Ford Fusion	83	0.367	14.16	10,227



* The centrality measures are with respect to all 135 car models in the network

Source: Netzer et al.

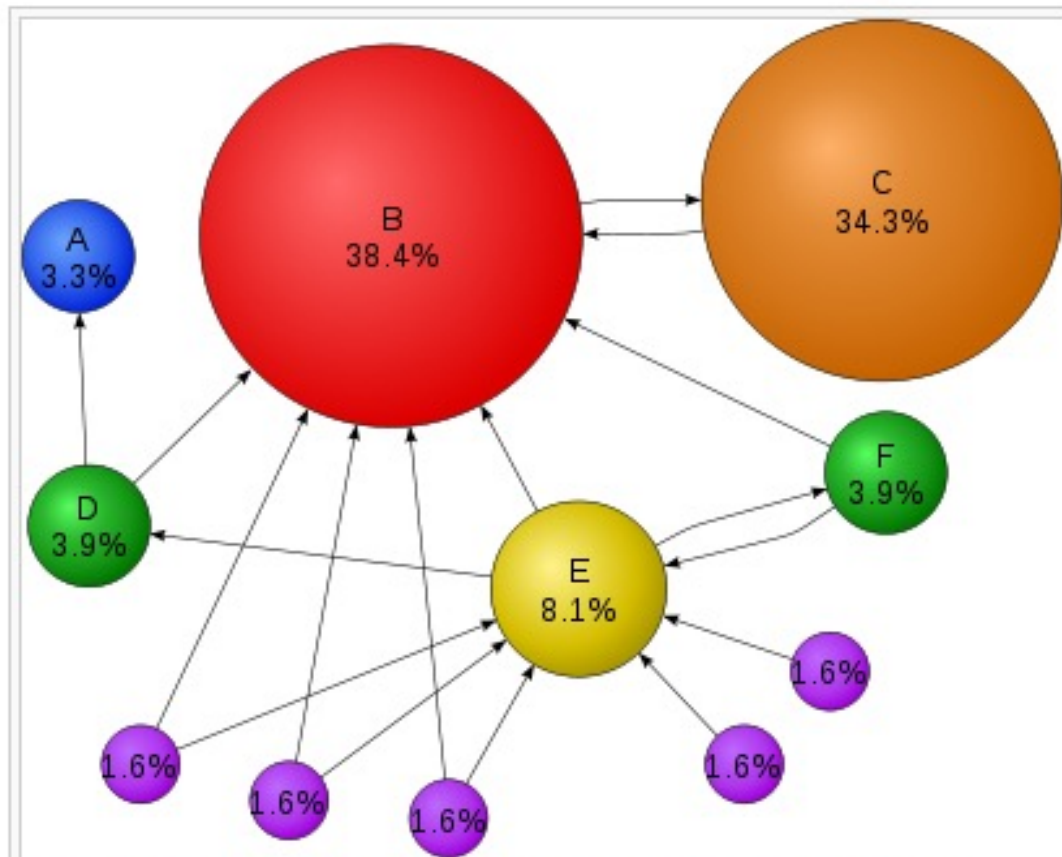


Metric 4: Eigenvector Centrality

- One's importance is partly determined by “the company one keeps”
- If one has many important friends, s/he should be important 😊
- Eigenvector centrality considers not only your degree, but your friends' degree
 - I.e., are your friends connected to large networks?
- Google's PageRank is closely related to Eigenvector centrality

PageRank of Web Pages

- If many web pages refer to a page, the latter must be important
- If the referring pages are referred to by many other pages, the effect is stronger



Mathematical **PageRanks** for a simple network, expressed as percentages. (Google uses a [logarithmic scale](#).) Page C has a higher PageRank than Page E, even though there are fewer links to C; the one link to C comes from an important page and hence is of high value. If web surfers who start on a random page have an 85% likelihood of choosing a random link from the page they are currently visiting, and a 15% likelihood of jumping to a page chosen at random from the entire web, they will reach Page E 8.1% of the time. (The 15% likelihood of jumping to an arbitrary page corresponds to a damping factor of 85%.) Without damping, all web surfers would eventually end up on Pages A, B, or C, and all other pages would have PageRank zero. In the presence of damping, Page A effectively links to all pages in the web, even though it has no outgoing links of its own.

Source: Wikipedia

The Original Google PageRank Algorithm



Web page x_1

Links to:

Page x_2

Page x_3

Page x_4

Web page x_3

Links to:

Page x_1

Web page x_2

Links to:

Page x_3

Page x_4

Web page x_4

Links to:

Page x_1

Page x_3

- As if a page transfers its importance to other pages
- E.g., x_1 transfers $1/3$ of its importance to each of x_2 , x_3 & x_4
- Graph representation, transition matrix, Eigenvector **PageRank** calculation
- Interpretation from a probabilistic perspective

The Network

