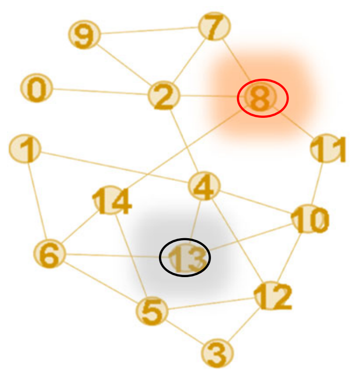
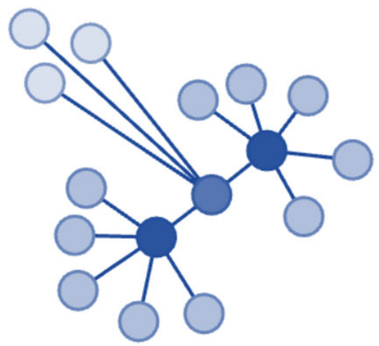
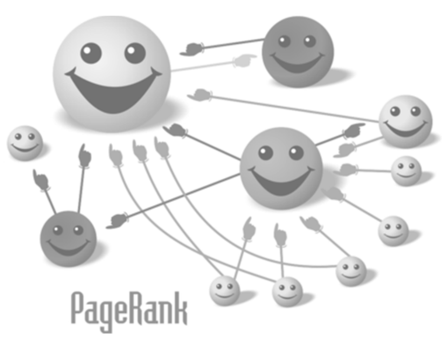
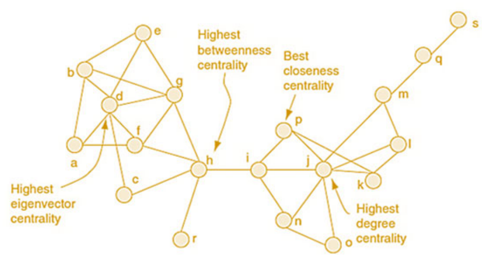
Prof. Ralucca Gera, rgera@nps.edu
Applied Mathematics Department,
Naval Postgraduate School









MA4404 Complex Networks

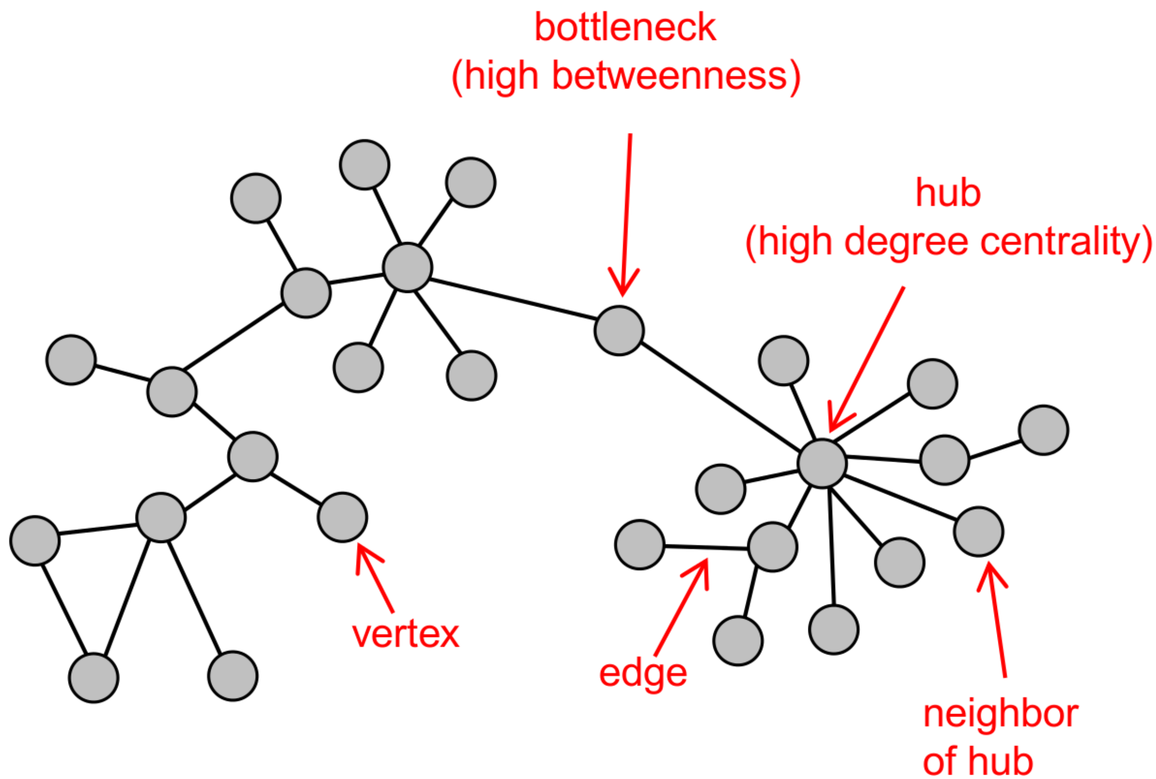
Betweenness Centrality

Learning Outcomes

- Compute Betweenness Centrality per node.
- Interpret the meaning of the values of Betweenness Centrality.

Why?!

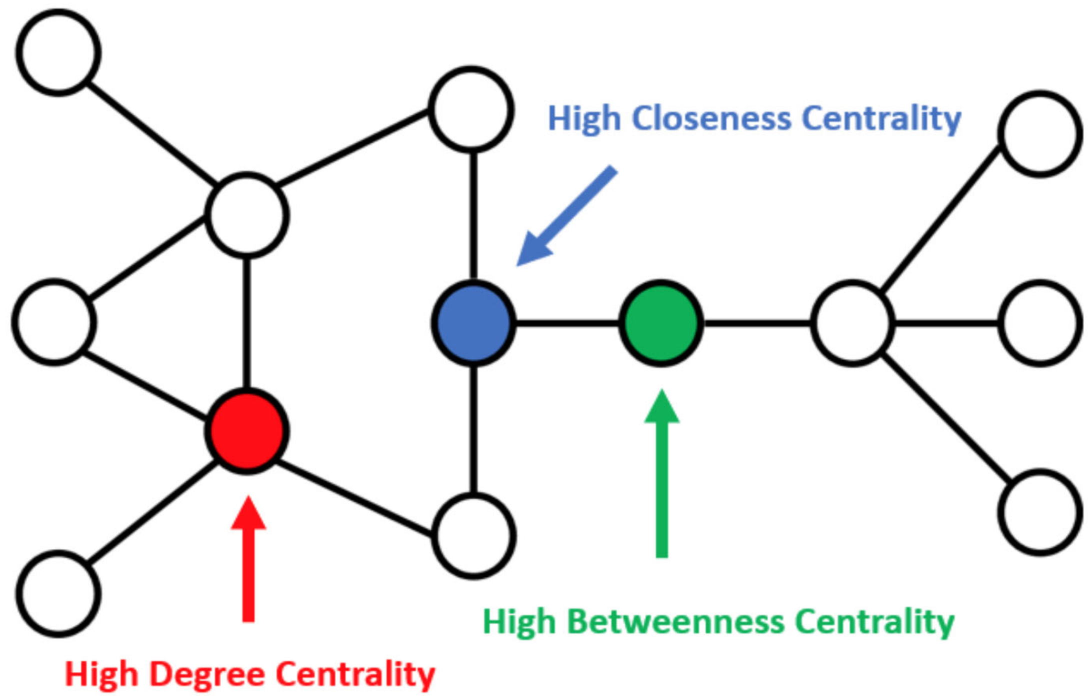
- **Intuition**: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?
- Interactions between two individuals depend on the other individuals in the set of nodes. The nodes in the middle have some control over the paths in the graph.
- Useful for flow, such as information or data packages



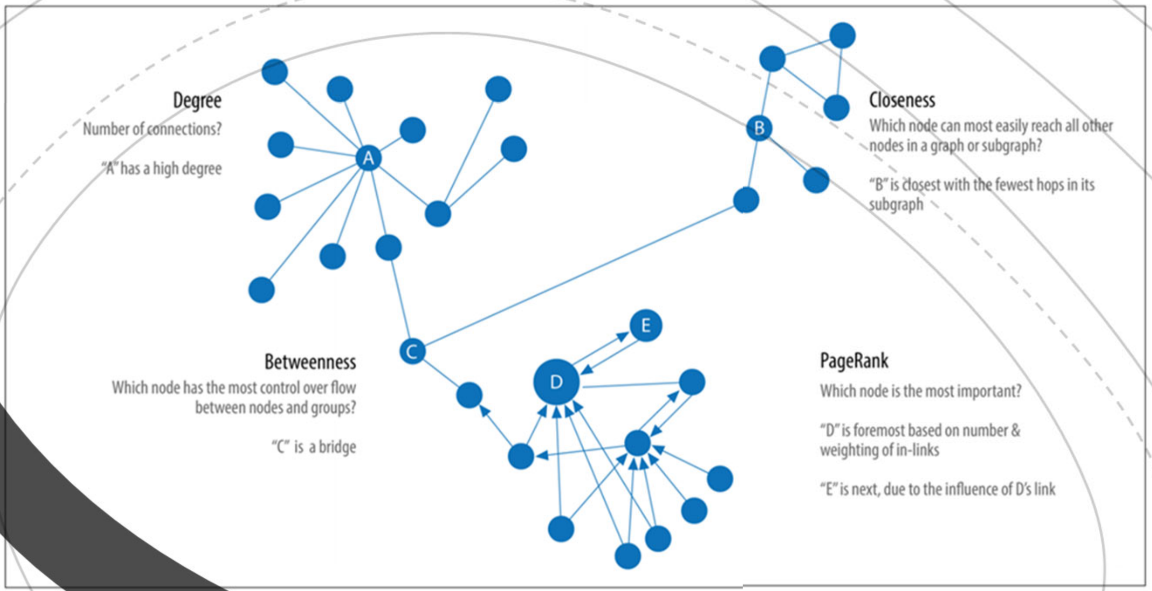
[PLOS ONE: A Network Integration Approach to Predict Conserved Regulators Related to Pathogenicity of Influenza and SARS-CoV Respiratory Viruses]

Assumptions

- If more than one geodesic → all geodesics are equally likely to be used.
- Flow takes the shortest path (we'll look at alternatives)
- Every pair of nodes exchanges a message with equal probability per unit time.
- Question: How many messages, on average, will have passed through each vertex en route to their destination?
  - A node's betweenness is given by all pairs of nodes, including the node in question.



https://essay.utwente.nl/76896/1/van%20Tilborg_MA_Faculty%20of%20Behavioural%2C%20Management%20and%20Social%20Sciences.pdf



Intuition Betweenness Centrality

Meaning of betweenness centrality

Vertices with high betweenness centrality have influence in the network by virtue of their control over information passing between others.

- They get to see the messages as they pass through
- They could get paid for passing the message along

Thus, their removal would disrupt communication

How would you capture it in a mathematical formula?

$$x_i = \sum n_{st}^i, \forall\ s, t \in V(G)$$

where

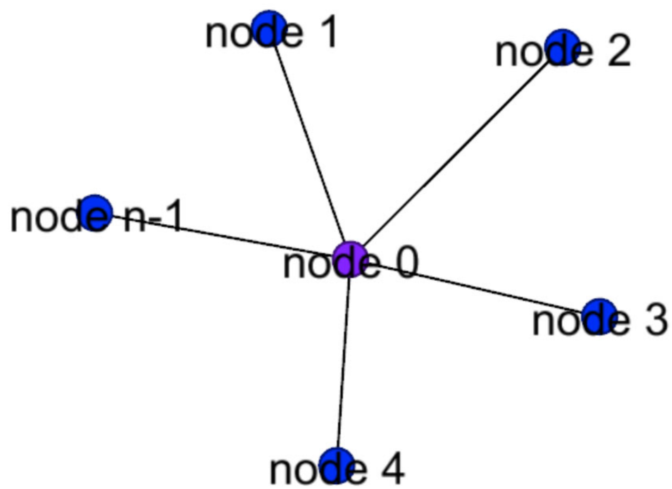
- n_{st}^i is the number of *s*-*t* geodesics that *i* belongs to (default: *i* cannot equal *s* or *t*)
- in an undirected graph, an *s*-*t* geodesic is the same as a *t*-*s* geodesics, so the edge gets counted twice)

It is applicable to directed networks as well.

Bounds for BC in connected graphs

Let G be a connected graph:

- What is the minimum value of betweenness centrality a vertex can have?
  - A leaf has: $(n-1)+(n-1)+1=2n-1$ since we have $n-1$ paths from x to each vertex, also $(n-1)$ more paths from each vertex to x, and one path from x to x.
- What is the maximum value of betweenness centrality a vertex can have?
  - The center of a star, say node 0: $n^2-(n-1)$ Let at $\mathrm{V(star)}=\{v, v_1, v_2, \dots, v_{n-1}\}$ with center node at v . Then there are n^2 pairs of nodes, from which we take away the $\mathrm{n}-1$ paths from each v_i to itself since $v_i \neq v$ (so v is not on any of these n-1)



A refined formula

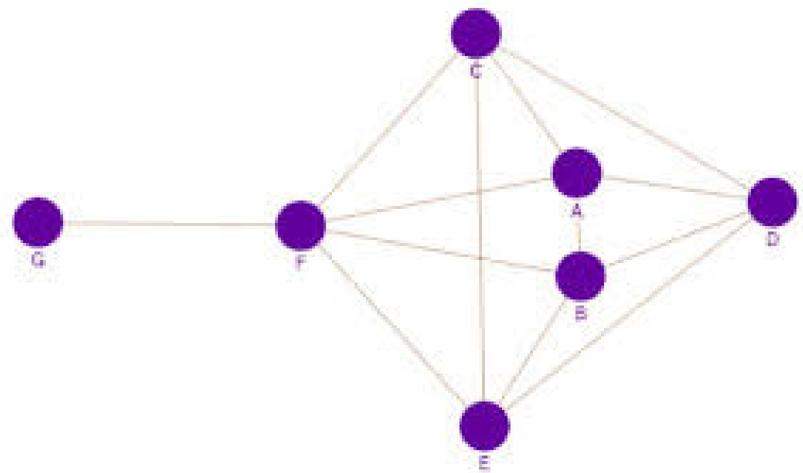
How do we find the relative (to the other nodes) betweenness centrality values?

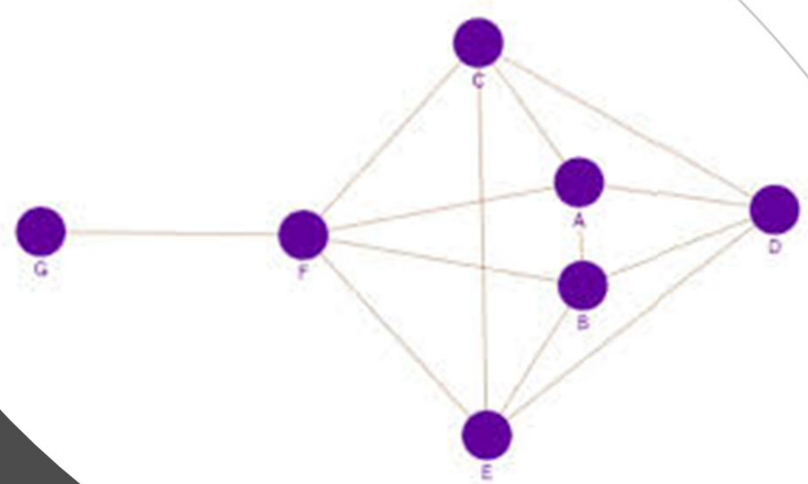
$$x_i = \sum \frac{n_{st}^i}{g_{st}}, \forall\, s,t \in V(G),$$

where:

- n_{st}^i is the number of *s-t* geodesics that *i* belongs to.
- g_{st} is the number of *s-t* geodesics
- Convention: if g_{st} = 0 and n_{st}^i = 0, then $x_i = 0$

(in an undirected graph, an *s-t* geodesic is the same as a *t-s* geodesics, so it gets counted twice)



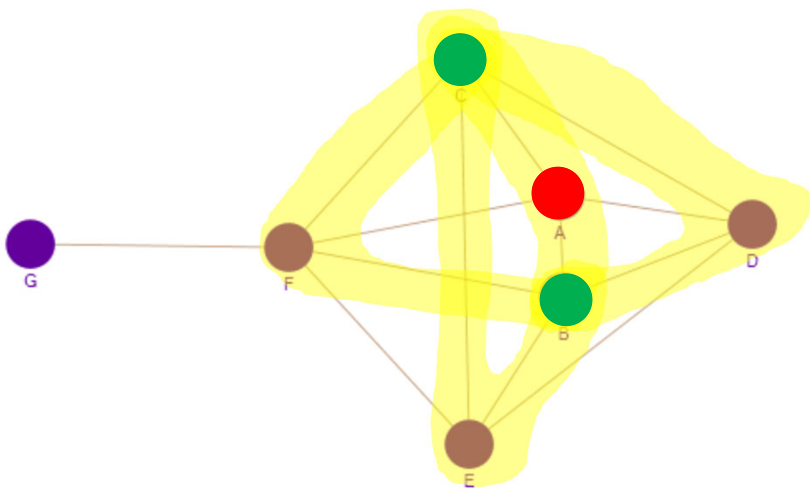


An in-class activity

What is the Betweenness of A?

Find the betweeneess of A: fraction of shortest paths that include vertex *A*

$$x_A = \sum \frac{n_{st}^A}{g_{st}}, \qquad \forall\, s,t \in V(G)$$



$$x_A = \sum \frac{n_{st}^A}{g_{st}} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 0.75$$

Number of paths

	A	B	C	D	E	F	G
A	-	1	1	1	4	1	1
B		-	4	1	1	1	1
C			-	1	1	1	1
D				-	1	4	4
E					-	1	1
F						-	1
G							

1 shortest path of 4 goes through A

1 shortest path of 4 goes through A

1 shortest path of 4 goes through A

Refinement of the BC formula

$$x_A = \sum \frac{n_{st}^A}{g_{st}}, \qquad \forall\, s,t \in V(G)$$

A normalized refined formula

How do we find the normalized relative betweenness centrality values? Allows to compare nodes in other graphs.

$$x_i = \left(\sum \frac{n_{st}^i}{g_{st}} \right) / \, n^2$$

where:

n_{st}^i is the number of *s*-*t* geodesics that *i* belongs to.

g_{st} is the number of *s*-*t* geodesics

Convention: g_{st} = 0 and n_{st}^i = 0, then $x_i = 0$

Another normalized formula

How do we find the normalized relative betweenness centrality values? Allows to compare nodes in other graphs.

$$x_i = \left(\sum \frac{n_{st}^i}{g_{st}} \right) / \left(n^2 - n + 1\right)$$

where:

n_{st}^i is the number of *s*-*t* geodesics that *i* belongs to.

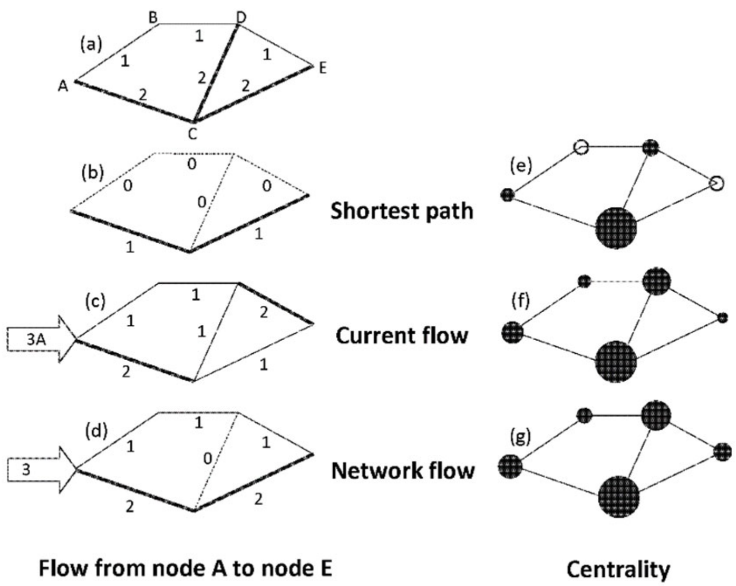
g_{st} is the number of *s*-*t* geodesics

Convention: g_{st} = 0 and n_{st}^i = 0, then $x_i = 0$

Extensions of Betweenness Centrality

- Used generally for Information flow
- Typically distributed over a wide range
- Betweenness only uses geodesic paths
- Information can also flow on longer paths
    - Sometimes we hear it through the grapevine

While betweenness focuses just on the geodesic, flow betweenness centrality focuses on how information might flow through many different paths, for example.



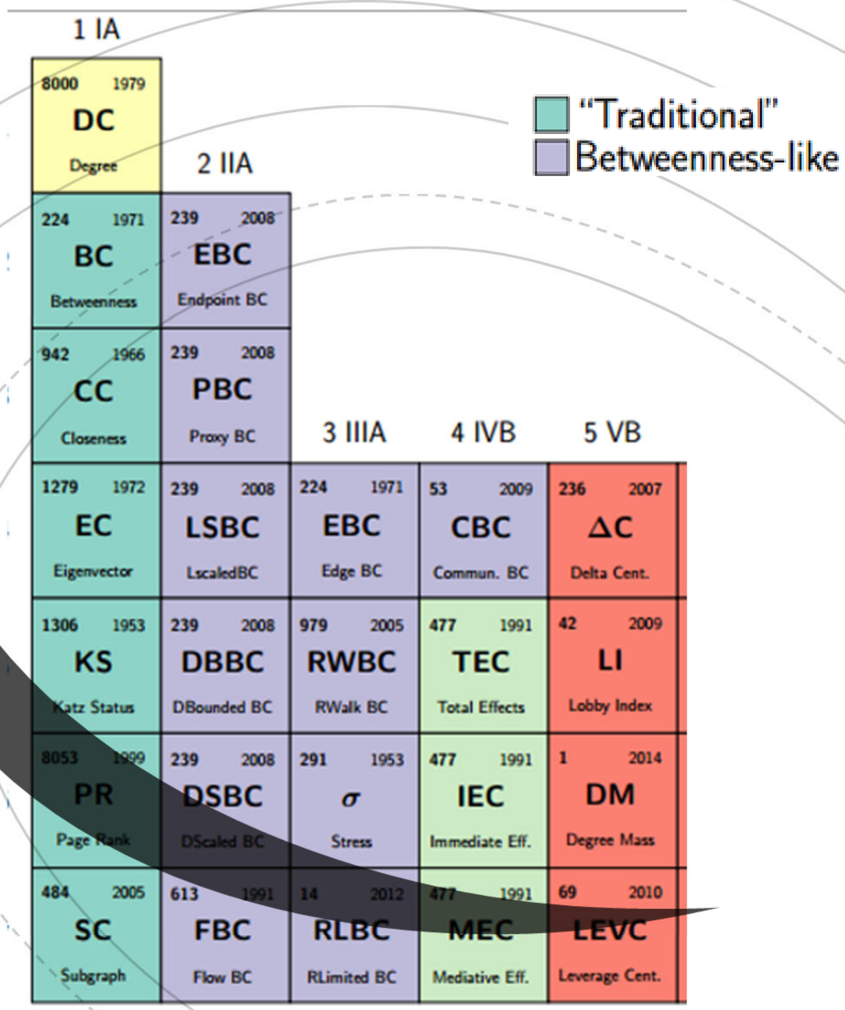
These methods consider respectively
1) the single shortest path,
2) probabilistic flow across all possible paths, and
3) optimal flow which considers, but may not use, all possible paths.

http://www.klamathconservation.org/CAT/v1_3_2/Help/module_6_1.htm

Extensions of Betweenness Centrality

- Used generally for Information flow
- Typically distributed over a wide range
- Betweenness only uses geodesic paths
- Information can also flow on longer paths
  - Sometimes we hear it through the grapevine

While betweenness focuses just on the geodesic, flow betweenness centrality focuses on how information might flow through many different paths, for example.



Extensions of Betweenness Centrality

Flow betweenness centrality

Same expression,

$$x_i = \sum \frac{n_{st}^i}{g_{st}}, \forall\, s, t \in V(G)$$

BUT

- n_{st}^i is the maximum flow transmitted from s to *t* through all possible paths that *i* belongs to.
- g_{st} is the maximum flow transmitted from s to *t* through all possible paths
- Convention: g_{st} = 0 and n_{st}^i = 0, then $x_i = 0$

(in an undirected graph, an *s*-*t* geodesic is the same as a *t*-*s* geodesics, so it gets counted twice)

Random walk betweenness centrality

Same expression,

$$x_i = \sum \frac{n_{st}^i}{g_{st}}, \forall\, s,t \in V(G)$$

BUT

- n_{st}^i is the number of times a random walk from *s* to *t* passes through *i*, averaged over many repetitions of a walk

- Note that $n_{st}^i \neq n_{ts}^i$
- A good measure for traffic that doesn't have a particular destination

Other extensions of centralities

- How would you extend the centralities you have seen? What else would you introduce that would capture the centrality of a vertex?
- Would you use it for edges?
- This is a good time to share your thoughts
- Subgraph/subset centrality?
  - How central are you to that particular subgraph?
  - How central is the subgraph to the network?
  - If so, would you repeat the centralities seen before for that subgraph?

A periodic table of centralities

Periodic Table of Network Centrality

	1 IA	2 IIA	3 IIIA	4 IVB	5 VB	6 VIB	7 VIIB	8 VIIIB	9 VIIIB	10 VIIIB	11 IB	12 IIB	13 IIIA	14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA
1	8000 1979 DC Degree																	518 1989 IC Information C
2	224 1971 BC Betweenness	239 2008 EBC Endpoint BC											26 1989 kPC kPath C.	275 2002 EGO Ego	51 2004 HYPER Hypergraphs	279 1997 AFF Affiliation C.	399 2 001 α-C α-Cent.	178 1995 ECC Eccentricity
3	942 1966 CC Closeness	239 2008 PBC Proxy BC											9068 1999 HITS Hubs/Authority	573 2006 g-kPC geodesic kPath	296 1999 GROUP Groups/Classes	80 2006 HYPSC Hyperg. SC	34 2010 t-SC t-Subgraph	116 1998 RAD Radiality
4	1279 1972 EC Eigenvector	239 2008 LSBC LscaledBC	224 1971 EBC Edge BC	53 2009 CBC Commun. BC	236 2007 ΔC Delta Cent.	5 2010 MDC MD Cent.	0 2015 EYC Entropy C.	2 2013 CAC Comm. Ability	56 2007 EPTC Entropy PC	281 1971 CCoef Clust. Coef.	42 2012 PeC PeC	427 2007 BN Bottleneck	43 2009 EI Essentiality I.	573 2006 e-kPC e-disjoint kPC	573 2006 v-kPC v-disjoint kPC	505 2010 WEIGHT Weighted C.	17 2013 TCom Total Comm.	116 1998 INT Integration
5	1306 1953 KS Katz Status	239 2008 DBBC DBounded BC	979 2005 RWBC RWalk BC	477 1991 TEC Total Effects	42 2009 LI Lobby Index	11 2008 MC Mod Cent.	0 2014 COMCC Community C.	45 2012 ECCoef ECCoef	0 2015 SMD Super Mediat.	1 2014 UCC United Comp.	4 2012 WDC WDC	119 2008 MNC MNC	43 2009 KL Clique Level	179 2005 BIP Bipartivity	426 1988 GPI GPI Power	116 1991 kRPC Reachability	58 2007 SCodd odd Subgraph	586 2004 RWCC RWalk CC
6	8053 1999 PR Page Rank	239 2008 DSBC DScaled BC	291 1953 σ Stress	477 1991 IEC Immediate Eff.	1 2014 DM Degree Mass	10 2012 LAPC Laplacian C.	0 2012 ABC Attentive BC	1699 2001 STRC Straightness C	0 2015 SNR Silent Node R.	15 2011 HPC Harm. Prot.	26 2011 LAC Local Average	119 2008 DMNC DMNC	3 2013 LR Lurker Rank	2457 1987 β-C β Cent.	X X HYP Hyperbolic C.	27 2012 kEPC k-edge PC	13 2007 FC Functional C.	0 2014 HCC Hierar. CC
7	484 2005 SC Subgraph	613 1991 FBC Flow BC	14 2012 RLBC RLimited BC	477 1991 MEC Mediative Eff.	69 2010 LEVC Leverage Cent.	35 2010 TC Topological C.	X X SDC Sphere Degree	15 2010 ZC Zonal Cent.	14 2013 CI Collab. Index	11 2013 CoEWC CoEWC	45 2012 NC NC	108 2010 MLC Moduland C.	X X RSC Resolvent SC	1 2014 SWIPD SWIPD	36 2009 XXXX LinComb	0 2014 BCPR BCPR	0 2014 TPC Tunable PC	0 2015 EDCC Effective Dist.

8000 1979 Freeman Conceptual	942 1966 Sabidussi Axiomatic	573 2006 Borgatti/Everett Conceptual	1130 2005 Borgatti Conceptual	24 2014 Boldi/Vigna Axiomatic	252 1974 Nieminen Axiomatic	6 1981 Kishi Axiomatic	3 2012 Kitti Axiomatic	3 2009 Garg Axiomatic
2065 1934 Moreno Historic	1546 1950 Bavelas Historic	780 1948 Bavelas Historic	1475 1951 Leavitt Historic	297 1992 Borgatti/Everett Conceptual	3649 2001 Jeong et al. Empirical	4167 1998 Tsai/Ghoshal Empirical	961 1993 Ibarra Empirical	71 2008 Valente Empirical

Key: citations year
C
Name

- "Traditional"
- Betweenness-like
- Friedkin Measures
- Miscellaneous
- Path-based
- Specific Network Type
- Spectral-based
- Closeness-like

http://schochastics.net/sna/periodic.html