November 19, 2011

A Novel Approach to Analyzing Online User Innovation Networks

Nitin Mayande, *Portland State University*
Charles Weber
Antonie J. Jetter, *Portland State University*

A Novel Approach to Analyzing Online User Innovation Networks*

Nitin Mayande, Charles Weber and Antonie Jetter

Portland State UNIVERSITY ETM

Mass Customization and Personalization Conference
San Francisco, California, USA
November 19, 2011

*This material is based upon work supported by the National Science Foundation under Grant No. 0822062 (Enabling Timely Revolutions in Organizational Performance). The authors would also like to acknowledge Intel Corporation for granting access to data upon which the research described in this paper is based.

Outline of Presentation

- MCPC: How granular is the multiverse?
- Extant theory of social networks
- Research questions
- Structure and information flow in online networks
- Metrics from graph theory
- Network archetypes
- Empirical study
- Simulation of an online network
- Implications and further research

Mass Customization and Personalization: How Granular is the Multiverse?

- A goal of mass customization is to design for the requirements of the individual user [12].
- But what if the user community consists an online network of millions of users?
- How many customizations do you realize?
- Do you design for the needs of the lead users [16]?
- Or the trendsetters in the mainstream [11]?
- How do you identify who is important?
- How do you make your product more valuable to your user community?

Key Gaps in Extant Theory

- ***Extant theory on social capital and social networks[1]***
  - ***Is primarily based on connectivity***
  - ***Does not take possibility of preferential attachments into consideration.***
    - Does not show how influence within a network depends upon control of information flow.
    - A critical factor contributing to influence within a network may have been ignored.
    - A manager's perceived influence may not reflect actual influence.
    - Managers may choose incorrect channels to get things done.

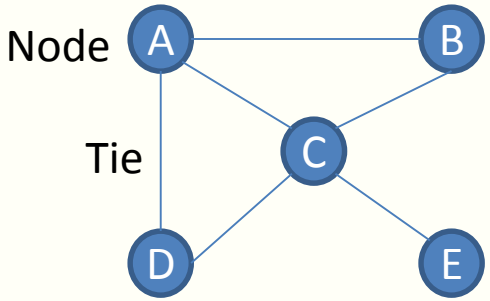
[1]e.g. Granovetter [10], Burt [5], Podolny [13], Coleman [6], Powell [14], Putnam [15]

Research Questions

- What are the best metrics and methods to determine the optimal locus of influence within a social network?
- How does one design a network that enhances or optimizes preferential attachments and information flow?
- How can the timing of surges in network performance be predicted in advance?
- Are there leading indicators?

About Network Structure and Information Flow

Borgatti [4]



- Human social networks can be represented as graphs.
- Nodes represent humans; ties represent relationships.
- Nodes that are directly connected by a tie are called neighbors.
- A ***path*** is a sequence of distinct nodes, with each node in the sequence being a neighbor of the preceding node.
- If information travels from the first node in the path to the last by following ties, then the number of ties that are traveled is the path's length.
- Multiple paths of varying lengths might lead from one node to another.
- The shortest path amongst such paths is called the ***geodesic***.

Serial Propagation (Borgatti, [4])

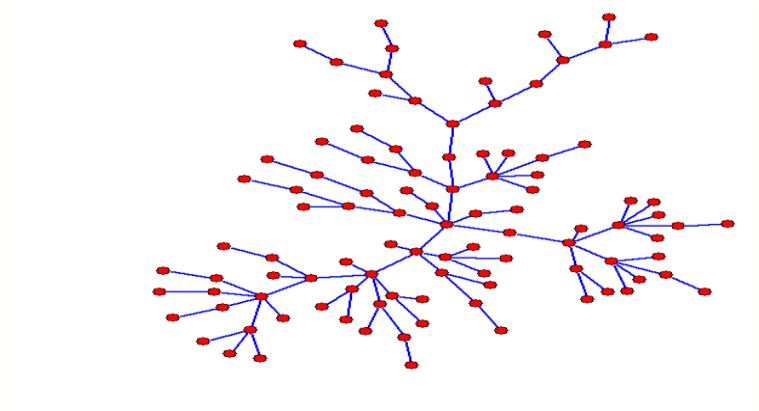
- Focus of this paper
- Propagation occurs by replicating what is at one node to multiple neighbors of the node one at a time.
- Example: gossip network amongst friends.
  - One person might pass the gossip to a friend, and then to another, and then to another.

Metrics for Measuring the Structure of Social Networks Based on Graph Theory

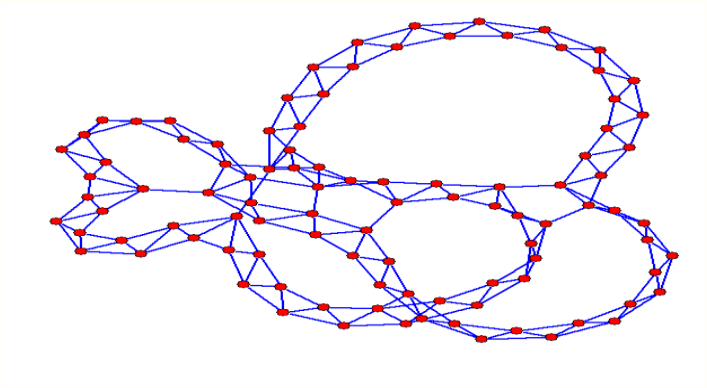
(Freeman, [7], [8])

- **Degree (communication activity)**:
  - The number of ties that are incident upon a node
- **Betweenness (communication control)**:
  - How often a node occurs on the shortest paths between other nodes?
- **Closeness (communication inefficiency)**:
  - Across how many nodes does the information have to travel to get from a given node to other reachable nodes in the network?
- **Eigenvector (importance of node)**: (Bonacich, [3])
  - connections to high-degree nodes contribute more to the score of the node in question than equal connections to low-degree nodes.
  - (Google's PageRank is a variant of the Eigenvector centrality measure.)

Network Archetypes



Scale Free Network (Albert & Barabási [1]; Goh *et al*. [9])



Small World Network (Watts & Strogatz [17])

Scale Free Networks:

- Exponentially distributed connections from each node outwards
- Uneven distribution means that some members are connected to a lesser / greater degree than others
- Greater degree of connection indicates a senior position in the network.
- *'Finding new friends on Facebook'*

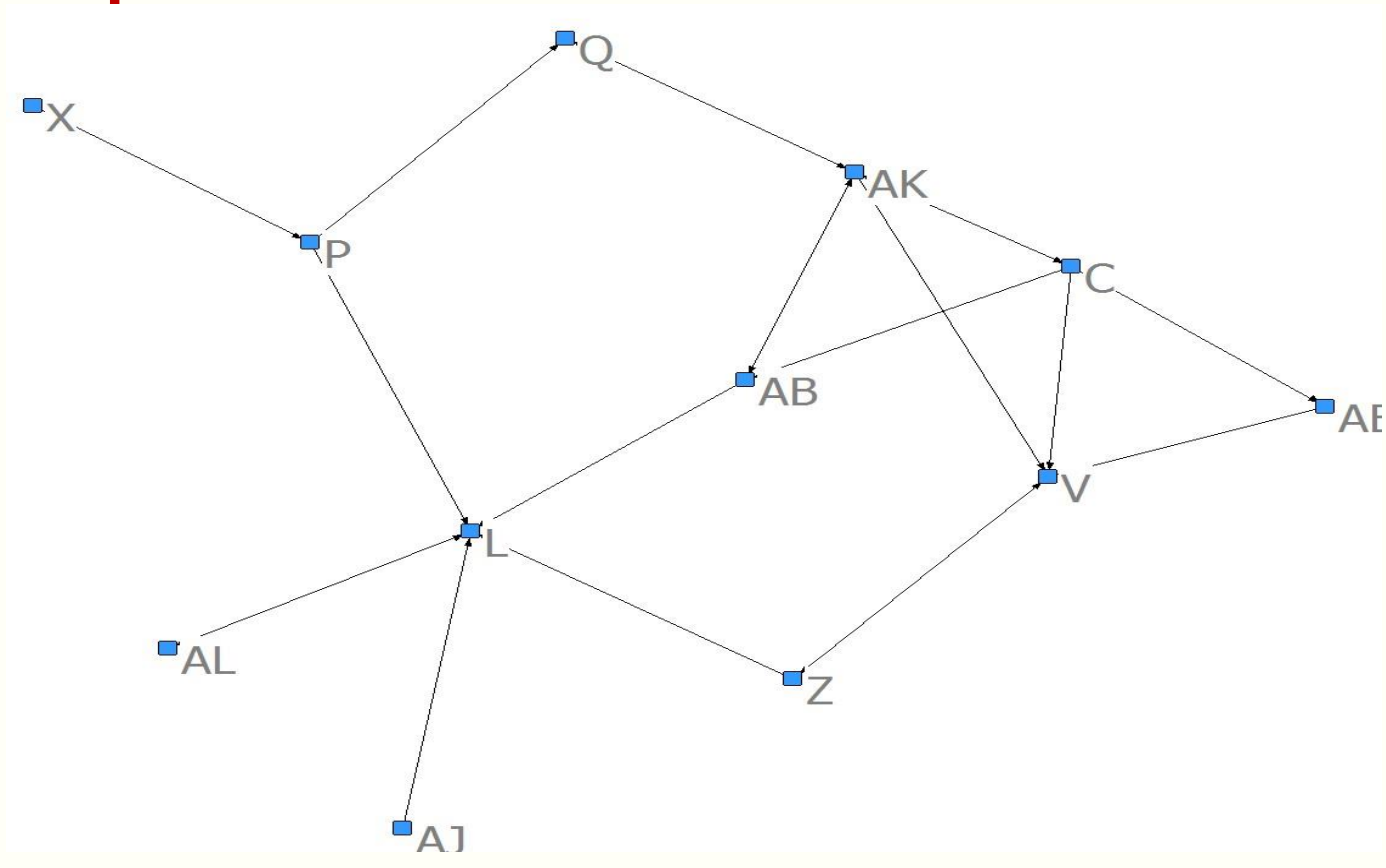
Small World Networks:

- Long distance connections are added at random to regular networks.
- Low path lengths between nodes
- High clustering coefficient (high levels of nodes' isolation)
- *'Reconnecting with old classmates on Facebook'*

Setting for Empirical Study:
the Intel Software Network

- For profit virtual network
- Purpose: to foster for profit peer production and open innovation within Intel's user communities.
- Ideal setting for study
  - Recent advances in Web2.0 technologies
    - are enabling for profit peer production and open innovation.
    - provide an elaborate record of the network's history including
      - network structure
      - information flow
  - Network exhibits hybrid structural characteristics.
    - Small world and scale free

Sample Network from Intel



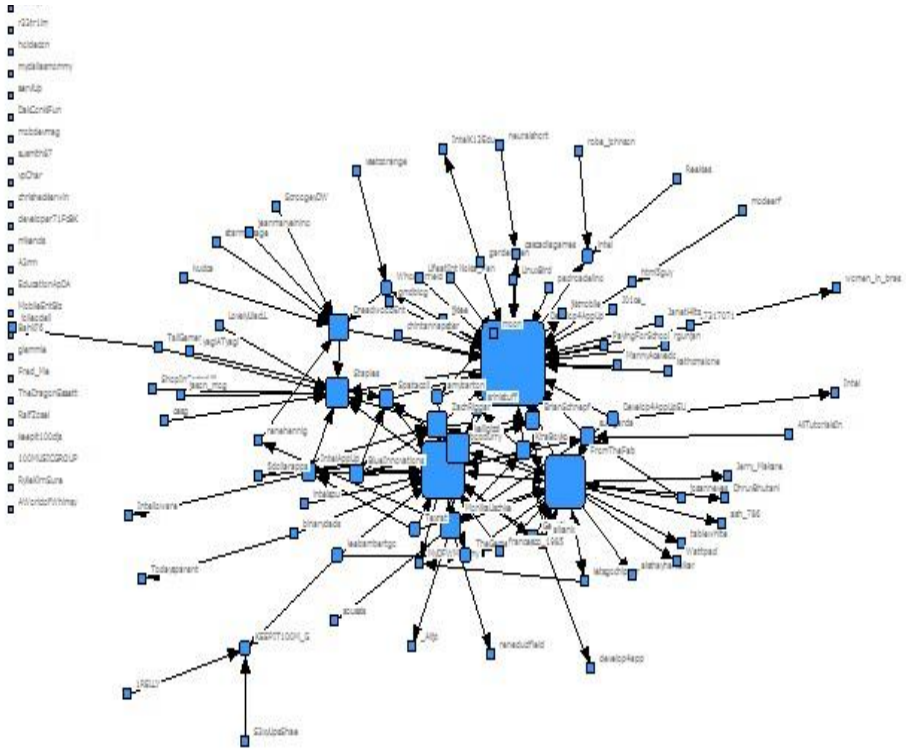
- Random sample of 12 nodes (users) and 15 ties (relationships)
- Serves as basis for comparing metrics

Comparing Measures of Influence

- All centrality measures seem to agree with each other to a certain extent.
  - For example, all centrality measures rank node **Q** around 8.
- The biggest disagreement in centrality measures occurs for node **C**.
  - Eigenvector centrality ranks node C at 1;
  - Degree centrality ranks it at 2-4;
  - Betweeness centrality ranks it at 7-8; and
  - Closeness centrality ranks it at 5-7.

Node	Degree Centrality	Between-ness Centrality	Closeness Centrality	Eigen-vector Centrality
AB	3.00 (5)	9.00 (3)	32.00 (2)	0.36 (4)
AE	2.00 (6-9)	0.00 (9-12)	43.00 (11)	0.27 (6)
AJ	1.00 (10-12)	0.00 (9-12)	40.00 (9-10)	0.09 (10-11)
AK	4.00 (2-4)	8.00 (4)	33.00 (3)	0.44 (2)
AL	1.00 (10-12)	0.00 (9-12)	40.00 (9-10)	0.09 (10-11)
C	4.00 (2-4)	5.00 (7-8)	35.00 (5-7)	0.46 (1)
L	5.00 (1)	27.50 (1)	30.00 (1)	0.28 (5)
P	3.00 (5)	13.50 (2)	35.00 (5-7)	0.16 (9)
Q	2.00 (6-9)	5.00 (7-8)	36.00 (8)	0.18 (8)
V	4.00 (2-4)	7.00 (5)	35.00 (5-7)	0.43 (3)
X	1.00 (10-12)	0.00 (9-12)	45.00 (12)	0.05 (12)
Z	2.00 (6-9)	6.00 (6)	34.00 (4)	0.22 (7)

Intel Software Network

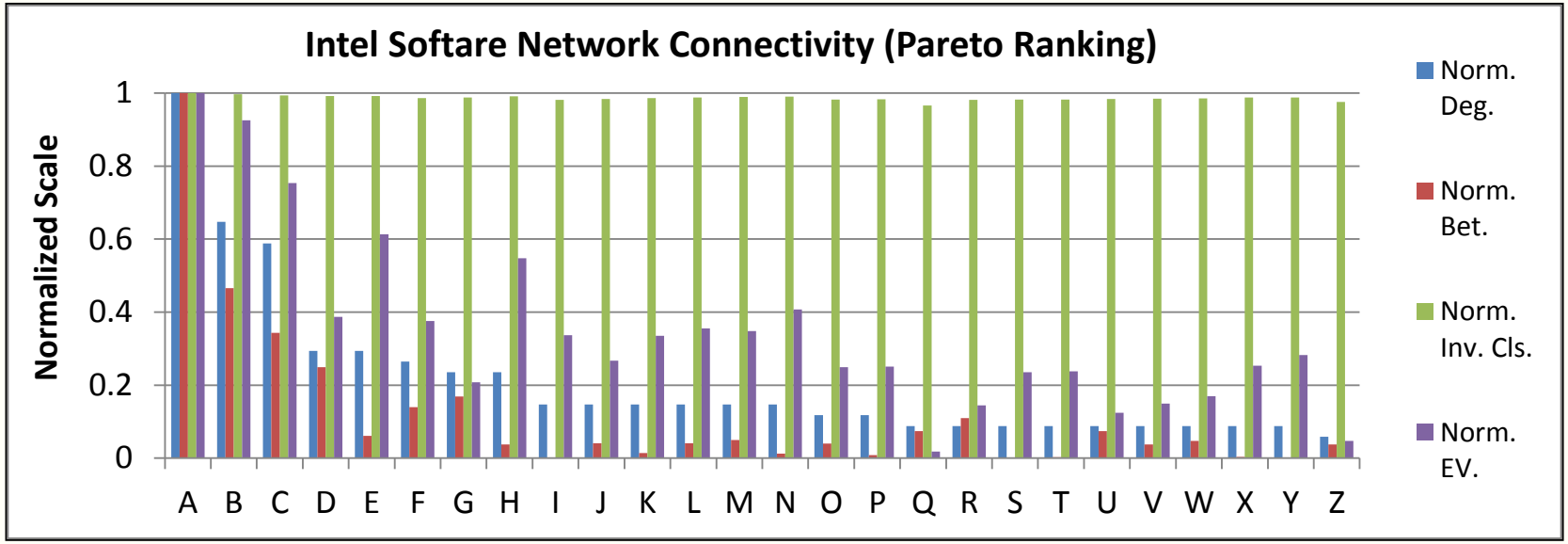


- Larger random sample
  - One week's Twitter traffic
  - 124 nodes
  - 89 connected
- Sparse sample
  - Total # of ties = 134
- More scale free than small world

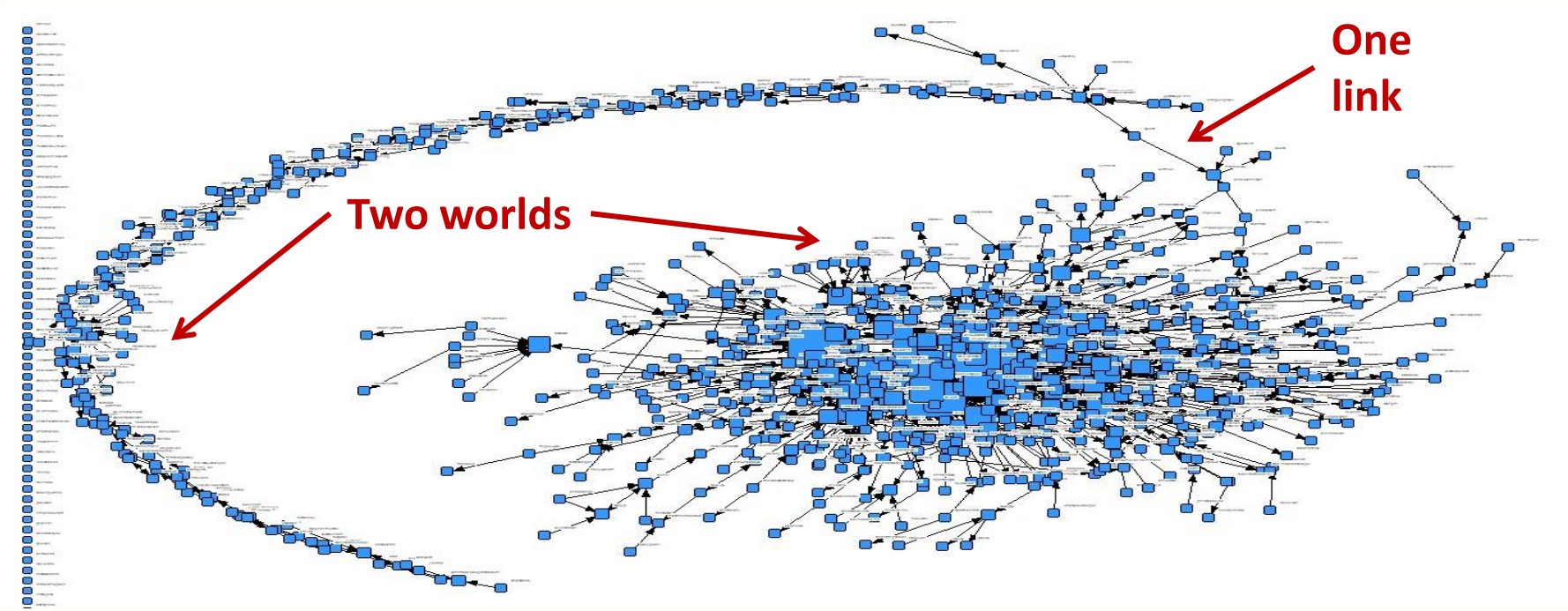
Connectivity of Intel Software Network

Intel Rank Correlation (Spearman's rho); N=124				
	Degree	Between-ness	Close-ness	Eigen-vector
Degree	1.000	0.787	-0.868	0.882
Betweenness	0.787	1.000	-0.647	0.671
Closeness	-0.868	-0.647	1.000	-0.964
Eigenvector	0.882	0.671	-0.964	1.000
Correlations are significant at the 0.01 level (2-tail)				

- Metrics disagree as connectivity decreases.
- Ranking varies from metric to metric.



VMWorld Software Conference

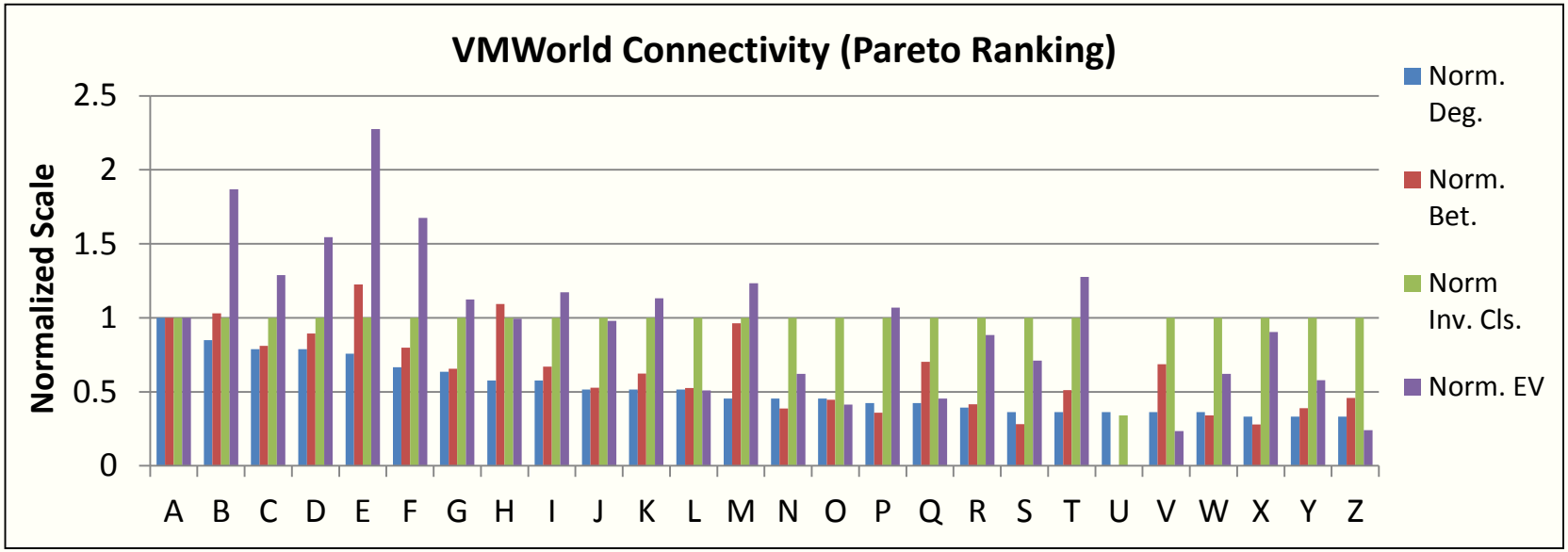


- Total Population: Two days' Twitter traffic;
  - 958 nodes; 868 connected
- Sparse sample -- # of ties =1044;
- More scale free than small world

Connectivity of the VMWorld Conference

VMWorld Rank Correlation (Spearman's rho); N=958				
	Degree	Between-ness	Close-ness	Eigen-vector
Degree	1.000	0.900	-0.706	0.618
Betweenness	0.900	1.000	-0.567	0.557
Closeness	-0.706	-0.567	1.000	-0.955
Eigenvector	0.618	0.557	-0.955	1.000
Correlations are significant at the 0.01 level (2-tail)				

- Metrics disagree as connectivity decreases.
- Ranking varies from metric to metric.



Tentative Conclusions and Questions

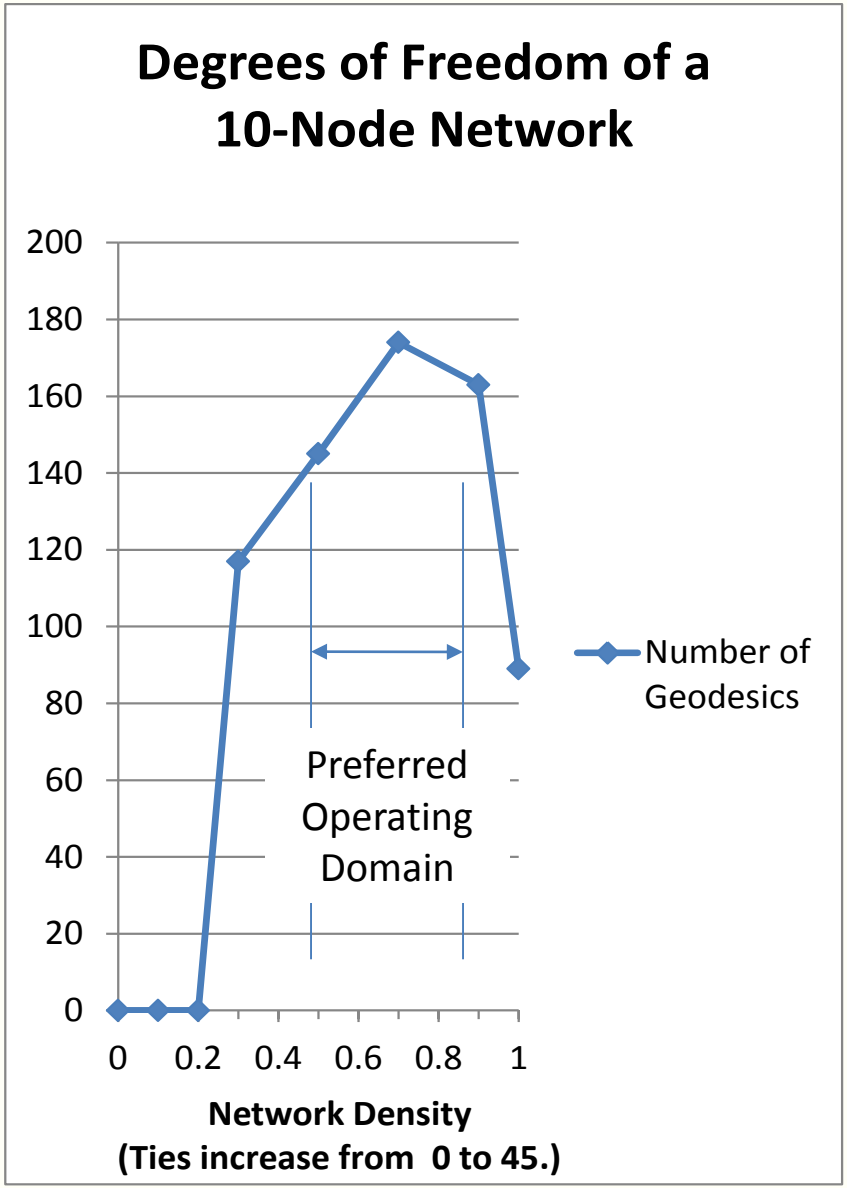
- The extent of influence within a social network may depend upon all of the above measures.
- To date no optimal measure for extent of influence has been identified.
- To identify true loci of influence, more research needs to be done.
- Specific questions:
  - Does the correlation between metrics decrease as network size increases?
  - What about network density?
  - What about information flow?

Simulation of a Small Network

- Number of geodesics as a function of network density
  - 10 nodes; up to 45 ties.
  - Randomly increase network density from 0 to 45/45.
  - *We assume freedom for information flow increases with number of geodesics.*

Finding : Optimal Operating Domain for Social Network

- Network too chaotic at low density.
  - Human activity below critical mass.
- Optimal domain for structure formation at intermediate level
  - Large number of geodesics
- Network performance deteriorates at high network density
  - Too many communication channels.
  - Fewer short communication channels
  - No 'small worlds'



Implications

- Human activity has no significant impact until ND hits critical mass.
- Surge in human activity follows.
- Activity peaks and deteriorates at high ND.
- Human activity within network is prerequisite to network performance.
- Surge in network performance likely to follow surge in human activity.
- ***Number of geodesics forecasts revolutions in network performance.***
- ***Intel is likely to know in advance when its network's performance will take off.***

Contributions and Work in Progress

- ***Contributions to date***
  - Demonstrate that current approaches are not well suited for determining the true loci of influence.
- ***Work in progress***
  - Developing methods of analysis that is based on preferential attachment and information flow.
  - Repeat benchmarking study

Upcoming Research

- Analyze structure of complete Intel network (>10,000 nodes)
  - Validate the current results of serial-geodesic flow
  - for undirected and unweighted (homogeneous tie strength) networks
  - with networks of multiple sizes and varying structure
  - (denser vs. less dense; scale free vs. small world).
- Content analysis
  - Analyze content of all messages sent within the Intel Software Network
  - Characterize context in which Intel's various user communities operate
  - Identify the interests of different user groups
- Apply approach to other settings
  - Validate the metrics for directed and weighted networks
  - Characterize the impact of preferential attachment
  - Assess the impact of various governance structures

List of References (1)

[1] Albert, R. and Barabási, A. L, "Topology of Evolving Networks: Local Events and Universality," *Physical Review Letters,* vol. 85, pp. 5234-5237, 2000.

[2] Allen, T., *Managing the Flow of Technology*, MIT Press, Cambridge, Mass., USA, 1984.

[3] Bonacich, P., "Factoring and weighting approaches to clique identification," *Journal of Mathematical Sociology*, vol. 2, pp. 113-120, 1972.

[4] Borgatti, S. P., "Centrality and Network Flow," *Social Networks*, vol. 27, pp. 55-71, 2005.

[5] Burt, R. S., *Structural Holes: The Social Structure of Competition,* Cambridge, Mass.: Harvard University Press, 1992.

[6] Coleman, J. S., "Social Capital in the Creation of Human Capital," *American Journal of Sociology,* vol. 94, S95-S120, 1988.

[7] Freeman, L. C ., "A set of measures of centrality based on betweenness," S*ociometry*, vol. 40, pp. 35-41, 1977.

[8] Freeman, L. C., "Centrality in social networks: Conceptual clarification," *Social Networks*, vol. 1, pp. 215-239, 1979.

[9] Goh, K.-I., Oh, E., Jeong, H., Kahng, B., and Kim, D., "Classification of scale-free networks", *Proc. Natl. Acad. Sci. USA,* vol. 99, pp. 12583–12588, 2002.

List of References (2)

[10] Granovetter, M., "The strength of weak ties," *American Journal of Sociology,* vol. 78, pp. 1360-1380, 1973.

[11] Moore, G., *Crossing the Chasm*, Harper Business Essentials, 1991.

[12] Pine, J. B., *Mass Customization: The Next Frontier of Business Competition*, Boston: Harvard Business School Press, 1993.

[13] Podolny, J. M., "A Status-Based Model of Market Competition," *American Journal of Sociology,* vol. 98, pp. 829–872, 1993.

[14] Powell, W. W. "Neither Market Nor Hierarchy: Network Forms of Organization", in *Research in Organization Behaviour,* Barry M. Staw and L.L Cummings (eds.), vol. 12, pp. 295-336, 1990.

[15] Putnam, R. D., "Bowling Alone: America's Declining Social Capital", *Journal of Democracy,* Vol. 6, No. 1, January, pp. 65-78, 1995.

[16] Von Hippel, E., *The Sources of Innovation*, New York, NY: Oxford University Press, 1988.

[17] Watts, D. J. and Strogatz, S. H., "Collective dynamics of 'small-world' networks," *Nature,* vol. 39, pp. 409-410, 1998.

About the Authors

- ***Nitin Mayande*** is an independent consultant in the area of virtual social network design. He received his training in communication engineering in India, and holds an MS in technology management from Portland State University, where he is pursuing his doctorate in the same field.

- ***Charles Weber*** received (among other degrees) a B.S. degree in engineering physics from the University of Colorado, Boulder; an M.S. degree in electrical engineering from the University of California, Davis; and a Ph.D. in management from MIT's Sloan School of Management. He joined Hewlett-Packard Company as a process engineer in an IC manufacturing facility. He subsequently transferred to HP's IC process development center, working in electron beam lithography, parametric testing, microelectronic test structures, clean room layout, and yield management. From 1996 to 1998, Charles managed the defect detection project at SEMATECH as an HP assignee. In December 2002, he joined the faculty of Portland State University where he is an associate professor of engineering and technology management.

- ***Antonie Jetter*** is an assistant professor of engineering and technology management at Portland State University. She received her diploma and PhD from the Rheinisch Westfaelische Technische Hochschule Aachen in 1998 and 2004 respectively. She has work experience as a manager of accounting, controlling and business planning in a high-tech start-up. Her research interests include open innovation and the fuzzy front end of new product development.