Social Footprints: Identifying the roles of nodes and links in massive social networks

Hayder Radha
Michigan State University

WAVES Lab
Outline

- How do we identify the “influence” of nodes in a network?
  - How do we identify the “influence” of nodes from structural information only (topology of a network)

- Can we identify more specific roles of individuals in complex social networks?
  - Influence models between leaders and followers
  - What type of leaders?

- How to measure the “importance” and “role” of links?
  - “Strength of weak links”
  - Closing the gap between macro- and micro- analysis
WAVES Lab: Research Areas

- **Signal sparsification and rank minimization**
  - NSF CCF
  - NSF CyberSEES

- **Communications and networking**
  - NSF TF
  - NSF NeTS

- **Signal processing of network graphs**
  - NSF Social Computing

- **Google (YouTube)**
Network Graphs
Network Graphs
Identifying “influential” friends

Stanford social network

Adamic, Adar WWW 2010
Graphs and adjacency matrices

Undirected Graph & Adjacency Matrix

Undirected Graph

Adjacency Matrix

http://www.stoimen.com/
Graphs and adjacency matrices

Weighted Directed Graph & Adjacency Matrix

Weighted Directed Graph

Adjacency Matrix

Michigan State  WAVES Lab  © 2014 Hayder Radha
“Influence" based on centrality measures

- **Degree Centrality**
  - Measures the immediate rate of spread of a replicable commodity by a node

- **Closeness Centrality**
  - *Average length of geodesic paths to all nodes in the network*

- **Betweenness Centrality**
  - The *number of geodesics on which a particular node lies*
Measuring the influence of a node in a social network

Eigenvector Centrality (EVC)

- A node’s “influence” is a function of its neighbors’ influence
- Recursive definition
- Does not assume shortest path flow
- Assumes an “influence process” for the diffusion of a commodity through the network
Eigenvector Centrality

Scalar form

\[ x(i) = \frac{1}{\lambda} \sum_{j \in \Gamma(v_j)} x(j) \]
\[ = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} x(j) \]

Vector form

\[ x = \frac{1}{\lambda} Ax \]
\[ \lambda x = Ax \]
EVC of 100 Node Barabasi-Albert Graph

- Node degree distribution follows a power law.
- In this drawing, node degrees go down as we move counter-clockwise on circle.
Limitations of Eigenvector Centrality

• EVC works well enough in graphs consisting of a single cluster/community of nodes.

• When a graph has polarity and contains multiple communities the principal eigenvector is “pulled” in the direction of the largest community, away from other, smaller communities.

• **Examples:**
  - Social graphs capturing competing ideas/views/ideologies
  - Wireless networks
  - Other graphs with high clustering coefficients
EVC of 100 Node Barabasi-Albert Graph

- Node degree distribution follows a power law.
- In this drawing, node degrees go down as we move counter-clockwise on circle.
The two subnets are copies of each other.
Network consist of 100 + 100 nodes.
EVC is able to identify the same nodes as “most central” in both networks.
The two subnets consist of 100 + 50 nodes. EVC assigns high centrality scores to nodes in the larger BA subnet, almost completely disregarding the smaller component.
New Centrality Measure Needed?

- When dealing with complex massive networks with a large number of clusters, we need to search and examine a **multi-dimensional vector space** (in the overall spectral space of the network graph)

- An “influential” node could have its energy concentrated in one or more of the dimensions of the multi-dimensional vector space
Eigenvector Centrality of Mesh Network
Principal Component Centrality (PCC)

- Measured using multiple eigenvectors in a P-dimensional spectral space of a graph
- A node’s PCC is the $\ell_2$ norm of its coordinates in the P-dimensional hyperspace formed by the P most significant eigenvectors as its basis.

Principal Component Centrality

Matrix Formulation

\[ C_P = \sqrt{((AX_{N \times P}) \circ (AX_{N \times P}))1_{P \times 1}} \]

\[ = \sqrt{(X_{N \times P} \circ X_{N \times P})(\Lambda_{P \times 1} \circ \Lambda_{P \times 1})} \]

• The Hadamard/ Schur/ entrywise product operator is used
Graphical Interpretation of PCC

- Spectral drawing of mesh graph in 3 dimensions
- Nodes are positioned based on first 3 eigenvectors.
- Nodes are colored according to $C_{15}$ (15 feature PCC).
$P=1$
$P = 3$
$P=5$
$P=10$
$P=15$
$P = 20$
$P=50$
$P=100$
$P=200$
Graphical Interpretation of PCC

- Spectral drawing of mesh graph in 3 dimensions
- Nodes are positioned based on first 3 eigenvectors.
- Nodes are colored according to $C_{15}$ (15 feature PCC).
Measuring influence using Principle Component Centrality (PCC)
How many "influential" nodes?

- What criteria should be used to choose an appropriate number of “features” for PCC?

- Time and space complexity of eigendecomposition is significant

- Prefer to compute PCC with fewer eigenvectors if possible
In N-dimensional hyperspace of centrality vectors,

- Compute phase angle between N-dimensional EVC and PCC vectors.
- Add another feature → recompute phase angle.

\[ \phi(P) = \arccos \left( \frac{C_P}{|C_P|} \cdot \frac{C_E}{|C_E|} \right) \]
Phase Angle PCC vs EVC Vectors
Graph’s adjacency matrix can be reconstructed using its constituents eigenvectors components.

Partial reconstruction can be attempted using subset of features.

\[
A_P = X_{N \times P} \Lambda_{P \times P} X_{P \times N}^T
\]
Graph Reconstructions
Related Problem Areas

- Can we identify more specific roles of individuals in massive social networks?
  - Leaders versus followers?
  - What types of leaders?

- What can be learned about the role of “individual” links among nodes?
  - How important each link to the overall network?
  - Can this be used for “denoising” massive networks?

- The interaction between users and content in multimedia social networks such as YouTube
Can we identify more specific roles of individuals in massive social networks?
- Leaders versus followers?
- What types of leaders?

What can be learned about the role of “individual” links among nodes?
- How important each link to the overall network?
- Can this be used for “denoising” massive networks?

The interaction between users and content in multimedia social networks such as YouTube
Identifying experts

Zhang, Ackerman, Adamic, WWW 2007

Portion of the Java Forum Q&A network
Leaders vs. Followers

- Leaders’ opinions are highly influential

- **Advertising** companies gain by giving free samples to “leaders” instead of a random population

- For **community health** campaigns, targeting interventions at community leaders have been shown to be more effective than applying them to random individuals

- For **administrative science**, identifying leaders results in effective product development teams with better work performance
Friedkin-Johnsen Influence Model

LUCI model for leaders and followers

- Outward interaction is influenced by external influences and own prior interactions
- Built on the Friedkin-Johnsen Influence Model

\[
y_i(t) = \rho_i(t) \sum_{\tau=1}^{\tau_{\text{max}}} \sum_{j=1}^{N} m_{j,i}(t - \tau) + \gamma_i(t) \sum_{\tau=1}^{\tau_{\text{max}}} \sum_{j=1}^{N} m_{i,j}(t - \tau) + e_i(t)
\]

**Outward interaction** = **External influence** + **Own (history) influence**
LUCI model for leaders and followers

- Outward interaction is influenced by external influences and own prior interactions
- Built on the Friedkin-Johnsen Influence Model

\[ y_i(t) = \rho_i(t) \sum_{\tau=1}^{\tau_{\text{max}}} \sum_{j=1}^{N} m_{j,i}(t - \tau) + \gamma_i(t) \sum_{\tau=1}^{\tau_{\text{max}}} \sum_{j=1}^{N} m_{i,j}(t - \tau) + e_i(t) \]
LUCI model for leaders and followers

• Outward interaction is influenced by external influences and own prior interactions

• Built on the Friedkin-Johnsen Influence Model

\[
\min_{\rho_i(t), \gamma_i(t)} \left( y_i(t) - \left( \rho_i(t) \sum_{\tau=1}^{\tau_{\text{max}}} \sum_{j=1}^{N} m_{j,i}(t - \tau) + \gamma_i(t) \sum_{\tau=1}^{\tau_{\text{max}}} \sum_{j=1}^{N} m_{i,j}(t - \tau) \right)^2 \right)
\]

Outward interaction = External influence + Own (history) influence
LUCI model for leaders and followers

- Outward interaction is influenced by external influences and own prior interactions
- Built on the Friedkin-Johnsen Influence Model

\[
y_i(t) = \sum_{j=1}^{N} m_{i,j}(t) = \rho_i \sum_{j=1}^{N} m_{j,i}(t-1) + \gamma_i \sum_{j=1}^{N} m_{i,j}(t-1) + e_i
\]

**Outward interaction** = **External influence** + **Own (history) influence**
Everything\textsuperscript{2}, or E2, “a collaborative Web-based community consisting of a database of interlinked user-submitted written material.”

In 2006: hosted by U. Michigan Ann Arbor. “We exist thanks to their generosity” (which is motivated by their academic curiosity, I suppose).”

E2 servers moved to MSU in 2007

**Validation on E2 data**
- 3.9 million interactions
- 8K users with 1500 “leaders”
- Area Under the Curve 90.3%
LUCI model for leaders and followers

Facebook data
- 3 million users
- 23 million edges
- Interaction data over one year
- Time sample \((t)\) is one month
LUCI model for leaders and followers

Facebook data
- 3 million users
- 23 million edges
- Interaction data over one year
- Time sample ($t$) is one month
LUCI model for leaders and followers

Facebook data
- 3 million users
- 23 million edges
- Interaction data over one year
- Time sample (t) is one month
Leaders versus non-leaders

Degree distributions for leaders are quite different from the degree distributions for non-leaders.
LUCI model for leaders and followers

Facebook data
- 3 million users
- 23 million edges
- Interaction data over one year
- Time sample \( (t) \) is one month
Neutrals are different

- **Neutrals**’ levels of interaction are independent of interaction levels of their friends
- Have the lowest average degree and are mostly connected to followers or other neutrals in the friendship graph
- The average eigenvector centrality of neutrals is two orders of magnitude lower than other user categories
LUCI model for leaders and followers

Facebook data
- 3 million users
- 23 million edges
- Interaction data over one year
- Time sample \((t)\) is one month

![Graph showing Extrovert Leaders, Introvert Leaders, Neutrals, and Followers with external and own history influence coefficients.](image)
“... 40% of executives describe themselves as introverts, including Microsoft’s Bill Gates, the über-investors Warren Buffett and Charles Schwab, ... .Odds are President Barack Obama is an innie as well. What does that mean? That introverts, not just extroverts, have the right stuff to lead organizations in a go-go, extroverted business culture”

Forbes: Why Introverts Can Make The Best Leaders, Nov. 2009, Jennifer B. Kahnweiler
Introvert Leaders

- “...Adam M. Grant, Francesca Gino, and David A. Hofmann conducted research that found some fallacy in the conventional wisdom, which is supported by years of academic research, that extroverts make the best leaders. They wrote in a Harvard Business Review article that their findings suggested that extroverts and introverts were equally successful in leadership roles overall, and that introverts, in certain situations, actually make better bosses.”

Introvert Leaders

The Introverted Leader
BUILDING ON YOUR QUIET STRENGTH
Jennifer B. Kahnweiler, Ph.D.

The Introvert’s Guide to Success in Business and Leadership
Lisa Petrilli

Michigan State WAVES Lab © 2014 Hayder Radha
Introvert Leaders

Introvert’s Road to Leadership

The Power of Being Quiet

Laurie Cain

Quiet

Influence

The Introvert’s Guide to Making a Difference

Jennifer B. Kahnweiler, PhD

Introverts Will Rule the World

William Wyatt

Michigan State

WAVES Lab

© 2014 Hayder Radha
Related Problem Areas

- Can we identify more specific roles of individuals in massive social networks?
  - Leaders versus followers?
  - What types of leaders?

- What can be learned about the role of “individual” links among nodes?
  - How important each link to the overall network?
  - Can this be used for “denoising” massive networks?

- The interaction between users and content in multimedia social networks such as YouTube
Importance of links in social graphs

- Granovetter’s seminal work “The strength of weak ties”
  - Importance of links that are perceived as “weak”
  - “Redundancy” of links that are perceived as “strong”
  - The paradox that exists between micro- and macro-level perception of social networks

- **Goal:** *quantify* the “strength of ties” in a topological sense that reflects social science theories, bridging micro- and macro-level views of social networks

Granovetter, “The strength of weak ties,” in American Journal of Sociology, 1973
Transitivity is a global measure for how “cohesive” or “redundant” a network is. The ratio of the number of triangles to the number of connected triples.
Transitivity matrix as a measure for the strength of individual links in a network

Transitivity Function

\[ \tau(W) = \frac{\alpha}{\beta} = \frac{\text{trace}[W^3]}{\text{trace}[W^2H]} \]

Transitivity Gradient

\[ \nabla_W \tau \triangleq \frac{\partial \tau}{\partial W} \]

Transitivity Matrix

\[ T = \nabla_W \tau \odot W \]

Transitivity Matrix: “Role” of individual links

\[ T = \frac{3}{\beta} W^2 \odot W - \frac{\alpha}{\beta^2} (WH + HW) \odot W \]

mutual neighbors combined degree
Transitivity Matrix: “Role” of individual links

Football network
Gradient of transitivity matrix and community detection
Gradient of transitivity matrix and community detection
Gradient of transitivity matrix and community detection

- Comparison with modularity based community detection
Gradient of transitivity matrix and community detection

- Comparison with modularity based community detection
Conclusions

- Multidimensional spectral analysis methods for massive graphs are more insightful than traditional approaches.
- New “Graph Transforms” can provide new insight into social networks, neural networks, sensor networks, etc.
- Aspects of signal processing, graph theory, information theory and machine learning can be integrated to develop new analysis tools for massive network graphs.
Role of edges in brain networks
(McGovern Institute for Brain Research, MIT)
On-going work: Youtube viewership analysis...

- Analysis of the flow of viewership using causality
More details can be found in....


- Muhammad U. Ilyas and Hayder Radha , "Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality," Proceedings of the IEEE International Conference on Communications (ICC’11), Kyoto, Japan, June 5-9, 2011.

Acknowledgement

- **Collaborator/Grad students/Postdocs**
  - Mohammad Aghagolzadeh
  - Iman Barjasteh
  - Yongju Cho
  - Chinh Dang
  - Ramin Eslami
  - Sauleh Etemadi
  - Yongying Gao
  - Mohammed Halloush
  - Rami Halloush
  - M. Usman Illyas
  - Shirish Karande
  - Syed Ali Khayam
  - KiMoon Lee
  - Jin Lee
  - Kiran Misra
  - Cliff Lampe/Jon Morgan
  - Nima Moghadam
  - Saad Qaisar
  - Alex Liu/M. Zubair Shafiq
  - Sohraab Soltani
  - Mingquan Wu

- **National Science Foundation**
- Google
- Yahoo
- Microsoft Research
- Kodak Research Lab
- ETRI (South Korea)
- Sharp Labs
- Technicolor
- State of Michigan - MEDC
- Michigan State Foundation
- MSU Graduate School