To Stay or Not to Stay: Modeling Engagement Dynamics in Social Graphs

Fragkiskos D. Malliaros[1] Michalis Vazirgiannis[1,2]

[1]École Polytechnique, France

[2]Athens University of Economics and Business, Greece

ACM International Conference on Information and Knowledge Management (CIKM 2013)

San Francisco, CA

October 27 – November 1, 2013

Outline

Outline

Social Media and Networks

- Online social networks and social media
- Easily accessible network data at large scale
- Opportunity to scale up observations
- Large amounts of data raise new questions

Objectives and Contributions

Modeling Engagement Dynamics

- Given a large social graph, how can we model and quantify the engagement properties of nodes?
- **User engagement** refers to the extend that an individual is encouraged to participate in the activities of a community
- Closely related property to the one of **node departure dynamics**
  - Similar to the decision of becoming member of a community, an individual may also decide to leave the network

Main Contributions

- Study the property of engagement and how it can be used for modeling the departure dynamics in social graphs
- Measures of engagement (node and graph level)
- Experiments: Properties and dynamics of real graphs
- Implications of our study on a new problem of robustness/vulnerability assessment

Objectives and Contributions
Modeling Engagement Dynamics

- Given a large social graph, how can we model and quantify the engagement properties of nodes?
- **User engagement** refers to the extend that an individual is encouraged to participate in the activities of a community
- Closely related property to the one of **node departure dynamics**
  - Similar to the decision of becoming member of a community, an individual may also decide to leave the network

Main Contributions

- Study the property of engagement and how it can be used for modeling the departure dynamics in social graphs
- Measures of engagement (node and graph level)
- Experiments: Properties and dynamics of real graphs
- Implications of our study on a new problem of robustness/vulnerability assessment

Outline

Problem Statement (1/2)

Goal:

Model and study the problem of node engagement in social graphs, from a **network-wise** point of view

- Consider information only about the underlying graph structure

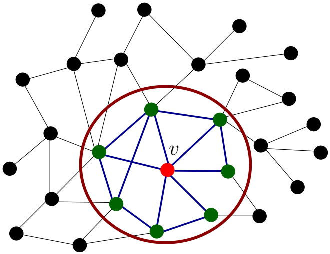
Problem Statement (1/2)

Goal:

Model and study the problem of node engagement in social graphs, from a **network-wise** point of view

- Consider information only about the underlying graph structure

- Each individual that participates in a social activity, derives a **benefit**
  - The benefit emanates from his/her neighborhood
- The benefit of each individual is affected by the degree of interaction among its neighbors **[Ugander et al., PNAS '12]**
  - If ones friends tend to highly interact among each other, the benefit of remaining engaged in the graph could potentially be increased

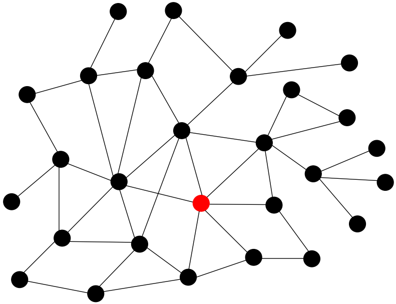


Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures

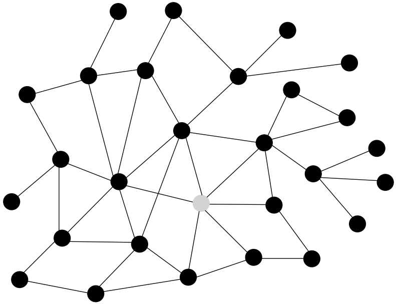
Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



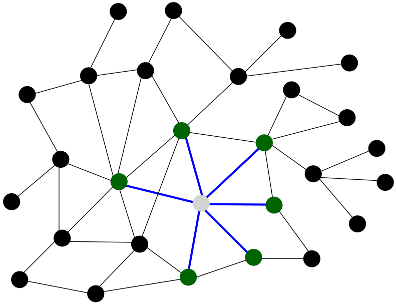
Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



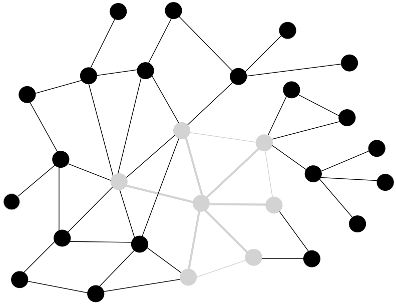
Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



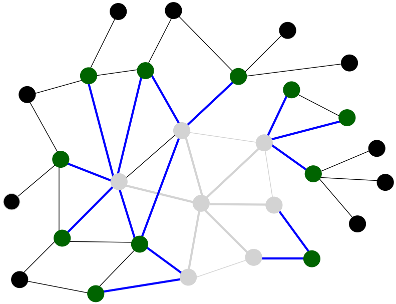
Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



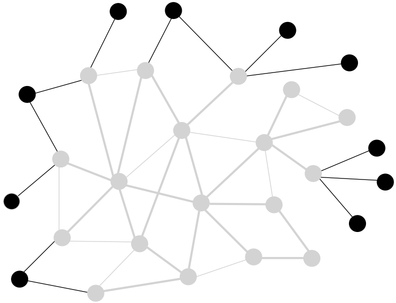
Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



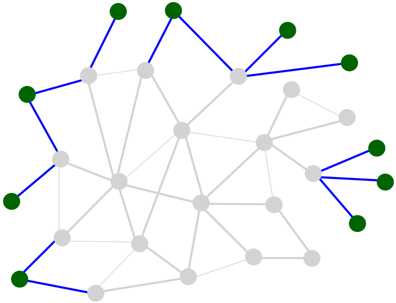
Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures

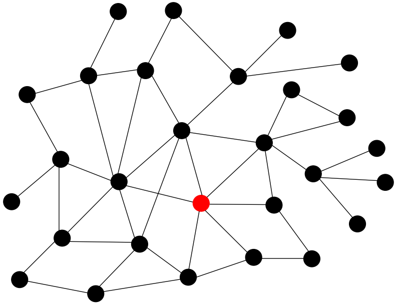


Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures

Problem Statement (2/2)

- Suppose now that a user decides to drop out due to the fact that the incentive of remaining engaged has been reduced
  - This decision will cause direct effects in his neighborhood $\longrightarrow$ Some of his friends may also decide to depart
  - A departure can become an **epidemic** (or contagion), forming a **cascade** of individual departures



- **Direct-benefit effects:** to incur an explicit benefit by remaining engaged, the decision of a node should align with the one of their neighbors **[Easley and Kleinberg, '10]**

Model Description (1/2)

- Each node $v \in V$ can either remain engaged or can decide to depart
- The behavior of nodes as a system can be captured by the notion of *networked coordination games* **[Easley and Kleinberg, '10]**
  - Network model based on direct benefit effects $\longrightarrow$ the benefit increases as more neighbors decide to stay
  - Nodes decide *simultaneously* whether to stay or leave

Model Description (1/2)

- Each node $v \in V$ can either remain engaged or can decide to depart
- The behavior of nodes as a system can be captured by the notion of *networked coordination games* **[Easley and Kleinberg, '10]**
  - Network model based on direct benefit effects $\longrightarrow$ the benefit increases as more neighbors decide to stay
  - Nodes decide *simultaneously* whether to stay or leave

Model Description (2/2)

- $\mathcal{X} = \{0, 1\}$: set of possible strategies (i.e., *leave* or *stay*)
  $\mathbf{x} = [x_1, x_2, \ldots, x_n]$: vector that denotes the decision of each node $i \in V$

- *Node payoff* function:
  $\Pi_i(\mathbf{x}) = \text{benefit}\left(x_i, \sum_{j \in \mathcal{N}_i} x_j\right) - \text{cost}(x_i), \mathcal{N}_i = \{j \in V : (i, j) \in E\}$
  - **Benefit function**: depends on node's own decision and the aggregate decision of the neighbors
  - **Cost function**: does not need to be known a priori $\rightarrow$ remain engaged if cost $\leq$ benefit (non-negative payoff)

Equilibrium Property [Manshadi and Johari, '09]; [Harkins, '13]

The best response of each node $i \in V$ corresponds to the *core number* c_i

Model Description (2/2)

- $\mathcal{X} = \{0, 1\}$: set of possible strategies (i.e., *leave* or *stay*)
  $\mathbf{x} = [x_1, x_2, \ldots, x_n]$: vector that denotes the decision of each node $i \in V$

- *Node payoff* function:
  $\Pi_i(\mathbf{x}) = \text{benefit}\left(x_i, \sum_{j \in \mathcal{N}_i} x_j\right) - \text{cost}(x_i), \mathcal{N}_i = \{j \in V : (i, j) \in E\}$
  - **Benefit function**: depends on node's own decision and the aggregate decision of the neighbors
  - **Cost function**: does not need to be known a priori $\rightarrow$ remain engaged if cost $\leq$ benefit (non-negative payoff)

Equilibrium Property [Manshadi and Johari, '09]; [Harkins, '13]

The best response of each node $i \in V$ corresponds to the *core number* c_i

Model Description (2/2)

- $\mathcal{X} = \{0, 1\}$: set of possible strategies (i.e., *leave* or *stay*)
  $\mathbf{x} = [x_1, x_2, \ldots, x_n]$: vector that denotes the decision of each node $i \in V$

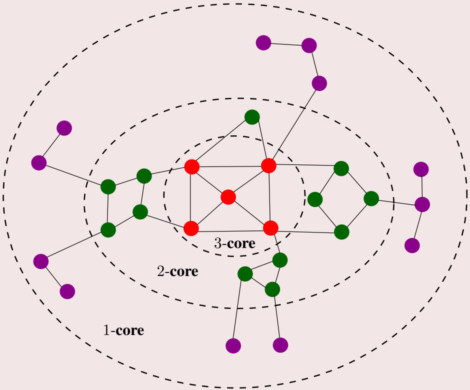
- *Node payoff* function:
  $\Pi_i(\mathbf{x}) = \text{benefit}\left(x_i, \sum_{j \in \mathcal{N}_i} x_j\right) - \text{cost}(x_i), \mathcal{N}_i = \{j \in V : (i, j) \in E\}$
  - **Benefit function**: depends on node's own decision and the aggregate decision of the neighbors
  - **Cost function**: does not need to be known a priori $\rightarrow$ remain engaged if cost $\leq$ benefit (non-negative payoff)

Equilibrium Property [Manshadi and Johari, '09]; [Harkins, '13]

The best response of each node $i \in V$ corresponds to the *core number* c_i

k-core Decomposition

Example



Core number $\mathbf{c_i} = \mathbf{1}$

Core number $\mathbf{c_i} = \mathbf{2}$

Core number $\mathbf{c_i} = \mathbf{3}$

Graph Degeneracy $\delta^*(\mathbf{G}) = \mathbf{3}$

Outline

Proposed Engagement Measures (1/3)

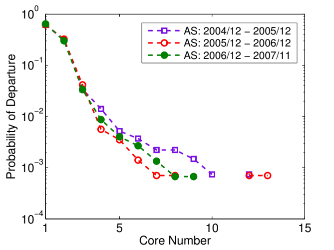
Node Engagement

Proposition (Node Engagement)

The engagement level e_i of each node $i \in V$ is defined as its core number c_i

- Nodes with higher core number $\longrightarrow$ better engagement

Prob. of departure vs. core number



CAIDA graph

- More refined modeling explanation of the **departure dynamics** in social graphs **[Wu et al., WSDM '13]**
  - **Active users:** core of the graph
  - **Inactive users:** periphery of the graph
  - The departure of nodes is *proportional* to their position in the graph

Proposed Engagement Measures (2/3)

Engagement Subgraphs

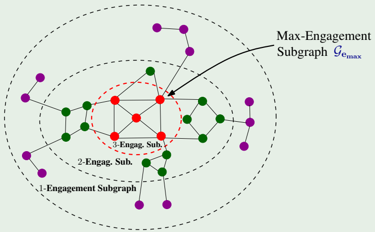
Definition (k-Engagement Subgraph $\mathcal{G}_k$)

The graph which is induced by the nodes $i \in V$ with engagement level $e_i \geq k$

Proposition (Max-Engagement Subgraph $\mathcal{G}_{e_{\max}}$)

- Let $k = \delta^*(G)$ be the degeneracy of the graph, i.e., the maximum k such that there exists a k-engagement subgraph
- Maximum engagement level of the graph: $e_{\max} = \delta^*(G)$
- Max-Engagement subgraph: composed by the nodes with engagement $e = e_{\max}$

Example graph



Proposed Engagement Measures (3/3)

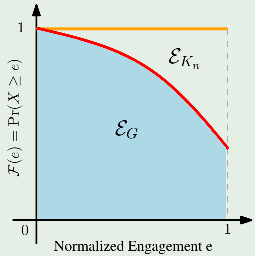
Graph Engagement

Definition (**Graph Engagement $\mathcal{E}_G$**)

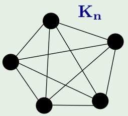
Let $\mathcal{F}(e) = \Pr(X \geq e)$ be the CDF of the sizes of the k-engagement subgraphs. Then, the total engagement level of a graph G, denoted as $\mathcal{E}_G$, is defined as the area under the curve of $\mathcal{F}(e)$, $e = [0, 1]$, i.e., $\mathcal{E}_G = \int_0^1 \mathcal{F}(e)\, de$

- Values in the range $[0, 1]$
- Higher $\mathcal{E}_G$ values $\rightarrow$ higher total engagement

Schematic representation of $\mathcal{E}_G$



Complete graph



Outline

Datasets

Basic Characteristics of Real-World Networks

Graph	# Nodes	# Edges
Facebook	63,392	816,886
YouTube	1,134,890	2,987,624
Slashdot	77,360	546,487
Epinions	75,877	405,739
Email-EuAll	224,832	340,795
Email-Enron	33,696	180,811
CA-GR-QC	4,158	13,428
CA-Astro-PH	17,903	197,031
CA-HEP-PH	11,204	117,649
CA-HEP-TH	8,638	24,827
CA-Cond-Mat	21,363	91,342
DBLP	404,892	1,422,263

Experimental Setup

Address the following points:

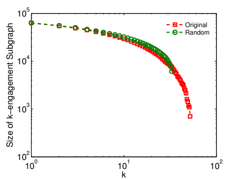
- P1 Study the characteristics of the engagement dynamics in real graphs
- P2 Examine how other graph features are related to the engagement of the graph

- Additional point: linear time complexity $\mathcal{O}(|E| + |V|)$
  - Properties of the k-core decomposition **[Batagelj and Zaversnik, '03]**

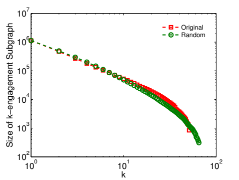
High Level Properties of k-Engagement Subgraphs (1/2)

Size Distribution

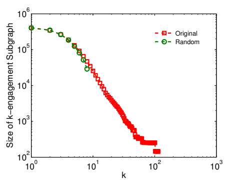
Size distribution of k-engagement subgraphs $\mathcal{G}_k$



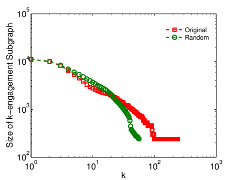
(a) FACEBOOK



(b) YOUTUBE



(c) DBLP



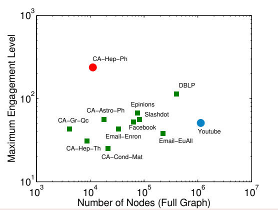
(d) CA-HEP-PH

- Almost **skewed** distribution
  - Small subgraphs with high engagement k
- Different behavior between **real** and **randomly** rewired graphs

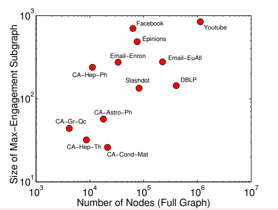
High Level Properties of k-Engagement Subgraphs (2/2)

Characteristics of Max-Engagement Subgraph $\mathcal{G}_{e_{max}}$

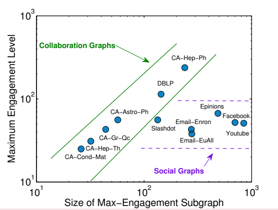
$|V|$ vs. e_{max}



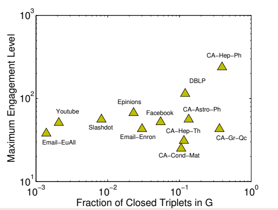
$|V|$ vs. # of nodes in $\mathcal{G}_{e_{max}}$



of nodes in $\mathcal{G}_{e_{max}}$ vs. e_{max}



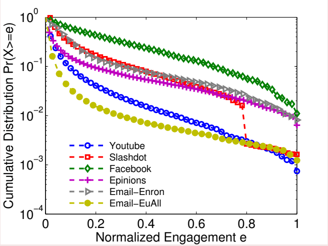
Closed triplets in G vs. e_{max}



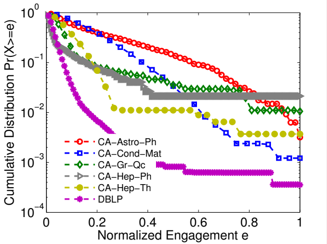
Graphs' Engagement Properties

Engagement Index $\mathcal{E}_G$

$\mathcal{E}_G$ index for social graphs



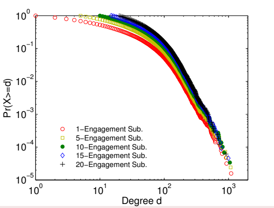
$\mathcal{E}_G$ index for collaboration graphs



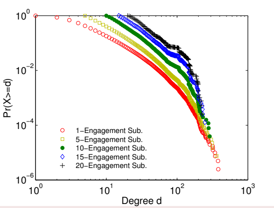
- FACEBOOK has the maximum engagement index $\mathcal{E}_G$
  - A relatively high fraction of nodes has high (normalized) engagement e
- DBLP shows the lower engagement index $\mathcal{E}_G$ in the collaboration graphs
  - **Possible explanation:** significant number of "relatively" new authors with low engagement

Near Self Similar *k*-Engagement Subgraphs

Cumulative degree distribution of $\mathcal{G}_k (k = 1, \ldots, 20)$



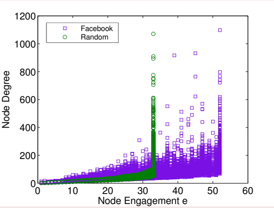
(a) FACEBOOK



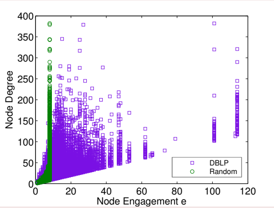
(b) DBLP

The shape of the cumulative degree distribution of $\mathcal{G}_k$'s is retained for various values of k

Node engagement *e* vs. node degree



(a) FACEBOOK

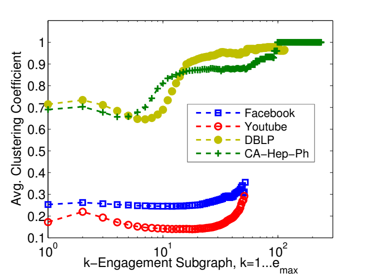


(b) DBLP

High degree nodes are possible to have low engagement

Engagement and Clustering Structures

Average clustering coefficient per $\mathcal{G}_k$



- Relation between engagement level and clustering structures in the graph
  - The probability of departure for a node is related to the overall neighborhood activity **[Wu et al., WSDM '13]**
- The avg. CC increases gradually as we are moving to $\mathcal{G}_k$'s of higher engagement

Outline

Disengagement Social Contagion

- **Robustness/vulnerability** assessment under node removals (departures) based on the engagement level
- The departure of a node can cause a **cascade** of node removals
  - We argue that nodes with high engagement will cause higher effect in the graph
- Almost **skewed** size distribution of the ***k***-engagement subgraphs for real-world graphs
  - **Random departures**
  - **Targeted departures**
- Robustness assessment similar to the seminal result by Albert, Jeong and Barabási **[Albert et al., Nature '00]**

Conclusions and Future Work

Contributions:

- Engagement property in social graphs and connection with the departure dynamics
- Measures of engagement at both node and graph level
- Experiments: Engagement dynamics of real graphs

Future work:

- Extend the study on more complex types of graphs (e.g., directed, signed)
- Robustness/vulnerability assessment under **targeted** and **random** node departures based on the engagement level

References I

D. Easley and J. Kleinberg
Networks, Crowds, and Markets: Reasoning About a Highly Connected World.
Cambridge University Press, New York, NY, USA, 2010.

V. H. Manshadi and R. Johari
Supermodular network games.
In *Allerton*, 2009.

A. Harkins
Network games with perfect complements.
Technical report, University of Warwick, 2013.

K. Bhawalkar, J. Kleinberg, K. Lewi, T. Roughgarden, and A. Sharma
Preventing unraveling in social networks: the anchored k-core problem.
In *ICALP*, 2011.

S. Wu, A. Das Sarma, A. Fabrikant, S. Lattanzi, and A. Tomkins
Arrival and departure dynamics in social networks.
In *WSDM*, 2013.

V. Batagelj and M. Zaversnik
An o(m) algorithm for cores decomposition of networks.
CoRR, 2003.

R. Albert, H. Jeong and A.-L. Barabási
Error and attack tolerance of complex networks.
Nature, 406, 378–382, 2000.

Acknowledgments

- **Google Europe Fellowship in Graph Mining**
  - Fragkiskos D. Malliaros
- **DIGITEO Chair Grant LEVETONE in France**
  - Michalis Vazirgiannis

Thank You !!

Fragkiskos D. Malliaros
Ph.D. Student
Data Science and Mining Group (DaSciM)
École Polytechnique, France
fmalliaros@lix.polytechnique.fr
www.lix.polytechnique.fr/∼fmalliaros

Michalis Vazirgiannis
Professor
Data Science and Mining Group (DaSciM)
École Polytechnique, France
mvazirg@lix.polytechnique.fr
www.lix.polytechnique.fr/∼mvazirg