The role of strong and weak ties in Facebook: a community structure perspective

Emilio Ferrara $^1,2$
with P. De Meo $^2$, G. Fiumara $^2$, A. Provetti $^2$

$^1$ Indiana University Bloomington – School of Informatics and Computing – CnetS – NaN
$^2$ Department of Physics, Informatics Section, University of Messina (IT)
Outline

1. Introduction
   - Strength of weak ties theory

2. Sampling and Analysis from Facebook
   - Sampling and building a Facebook dataset
   - Facebook network analysis

3. Strength of weak ties theory evaluation
   - Nature of ties in Facebook

4. Conclusion
   - Conclusions, open problems and future work
1. Introduction
   - Strength of weak ties theory

2. Sampling and Analysis from Facebook
   - Sampling and building a Facebook dataset
   - Facebook network analysis

3. Strength of weak ties theory evaluation
   - Nature of ties in Facebook

4. Conclusion
   - Conclusions, open problems and future work
In 1973, Mark Granovetter proposed the **strength of weak ties** theory: human relationships (acquaintance, loose friendship) that are *less binding* than family and close friendship but yield *better access to information and opportunities.*

He formalized the concept of **tie strength** in human relations as a (probably a linear) combination of **amount of time**, emotional intensity, intimacy (mutual confiding) and the reciprocal services which characterize the tie.

He also defined the most ideal context of assessment of this theory:

Ties are assumed to be **positive and symmetric**. Discussion of operational measures and weights is postponed to future empirical studies.

---

Strength of weak ties - operationalization

Requirement: a Social Network representable by means of mutual unweighted friendship relations. Facebook is probably the best-known example of such a type of friendship connection.

[Strict] **Weak ties** act as *bridges* among communities otherwise disconnected.

[Relaxed] **Weak ties** act as *shortcut bridges* shortening distances among nodes belonging to different communities.

We define, in our context of Facebook:

- **Weak Ties** Those edges that occur among nodes belonging to different communities.
- **Strong Ties** The vice-versa, those edges occurring among nodes in the same community.
Strength of weak ties - ideas

Scope  To investigate the **nature of ties** in Facebook to prove their **role** in the Social Networks.

We believe this an important task, since tie strength is proven to affect:

- **information diffusion** [Zhao] $^2$
- diffusion of **opinions and social influence** [Grabowicz] $^3$
- identification of **leaders** in the network [Centola] $^4$

Plan  To assess the **strength of weak ties theory** in Facebook we will:

1. Partition the network (i.e., community detection)
2. Classify links (weak ties - strong ties)
3. Study their features

---

$^2$Weak ties: A subtle role in the information diffusion of OSNs. 2010

$^3$Social features of online networks: The strength of intermediary ties in online social media, PLoS ONE 2012.

$^4$The spread of behavior in an online social network experiment, Science 2010.
Actors, ties and communities in Facebook in a shot!

The role of strong and weak ties in Facebook
1. Introduction
   - Strength of weak ties theory

2. Sampling and Analysis from Facebook
   - Sampling and building a Facebook dataset
   - Facebook network analysis

3. Strength of weak ties theory evaluation
   - Nature of ties in Facebook

4. Conclusion
   - Conclusions, open problems and future work
Sampling from Facebook: MHRW vs. Uniform sampling

Metropolis-Hastings Random Walk features:

- Biased towards low-degree nodes
- Transition probability given by

\[ P_{u,v}^{MHRW} = \frac{\min(1, k_v / k_w)}{k_v} \text{ if } w \text{ neighbor of } v \]

\[ P_{u,v}^{MHRW} = 1 - \sum_{y \neq v} P_{u,y}^{MHRW} \text{ if } u = v \]

Uniform (rejection-based) sampling: a list of random nodes to be visited is generated.

- Independent w.r.t. the structural distribution of friendship ties
- Produces unbiased results

---

\(^5\)By Gjoka et al. Walking in Facebook: A Case Study of Unbiased Sampling of OSNs, 2010
MHRW & Uniform dataset (acquired by Gjoka\textsuperscript{6} during August 2009)

<table>
<thead>
<tr>
<th></th>
<th>No. nodes</th>
<th>No. edges</th>
<th>Avg. deg.</th>
<th>Clust. coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHRW</td>
<td>957K</td>
<td>58.4M</td>
<td>94.1</td>
<td>0.05</td>
</tr>
<tr>
<td>Uniform</td>
<td>984K</td>
<td>72.2M</td>
<td>95.2</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Similar degree distribution showed by MHRW and Uniform samples

Do MHRW and Uniform complement each other? Can be merged?

\textsuperscript{6}Walking in Facebook: A Case Study of Unbiased Sampling of OSNs, 2010
Is it safe to merge the two networks? i.e., does merging introduce bias? Gjoka et. al\(^7\) proved that MHRW preserves statistical features and network structure. E.g., degree distributions of MHRW and Uniform networks overlap

\(^7\)Walking in Facebook: A Case Study of Unbiased Sampling of OSNs, 2010
The network exhibits two types of nodes:

1. Nodes visited during the sampling process;
2. Nodes only *seen* by the sampling as neighbors of other visited ones.

**Consideration**

To ensure reflecting the community structure, we retain only those nodes completely visited, and not only discovered by the sampling process.

In the pruned network we:

- Deleted all nodes which have only been *discovered* (i.e., the *frontier*),
- Deleted all edges pointing to frontier nodes.

<table>
<thead>
<tr>
<th></th>
<th>No. nodes</th>
<th>No. edges</th>
<th>Avg. deg.</th>
<th>Clust. coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SocialGraph</td>
<td>613K</td>
<td>2.04M</td>
<td>22.74</td>
<td>0.18</td>
</tr>
</tbody>
</table>
We collected geographical information about users in the sample to further study influence of physical distances on tie strength

- Geocoding Regional networks identifiers
- Yahoo PlaceFinder API
Outline

1. Introduction
   - Strength of weak ties theory

2. Sampling and Analysis from Facebook
   - Sampling and building a Facebook dataset
   - Facebook network analysis

3. Strength of weak ties theory evaluation
   - Nature of ties in Facebook

4. Conclusion
   - Conclusions, open problems and future work
In OSNs degree distribution usually follows a **power law** $P(k) \sim k^{-\gamma}$, with $\gamma \leq 3$.

There exists a small number of nodes with high degree (so-called **hubs**) and a large number of nodes with few connections (**tail of the power law**).

We plot the **Complementary Cumulative Distribution Function (CCDF)** $\varphi(k) = \int_k^\infty P(k') dk' \sim k^{-\gamma} \sim k^{-(\gamma-1)}$ in a **log-log scale**.

In our case we obtain $\gamma = 2.45$ (MEL fitting, $p$-value $< 0.001$).
Facebook community structure

**Community** Communities formation in SNs is a natural process since individuals aggregate together in groups that represent friendship, kinship, working relations etc.

**Definition** The so-called community structure is defined by an increased density of connections among nodes belonging to a given community with respect to nodes outside the community.

**Formulation** Commonly, we define the problem of finding a reasonable partitioning $V = (V_1 \cup V_2 \cup \cdots \cup V_n)$ of disjoint subset of nodes of a social graph $G = (V, E)$.

**Algorithms** A famous paradigm proposed by [Newman]$^8$, maximizes a function called network modularity $Q = \sum_{s=1}^{m} \left[ \frac{l_s}{|E|} - \left( \frac{d_s}{2|E|} \right)^2 \right]$ where $l_s$ is the number of edges belonging to the $s$-th community and $d_s$ is the sum of the degree of these nodes.

**Remark** High values of $Q \in [-0.5, 1]$ imply a clear emergence of the community structure.

---

Different modularity maximization strategies exist: we adopt the well-known *Louvain method* [Blondel] \(^9\)

Modularity based methods may suffer from a resolution limit [Fortunato] \(^10\), i.e. the inability of finding communities smaller than \(\sqrt{E/2} \approx 1,000\) in our case

We describe the community size CCDF and overall statistics on clustering

---

No. communities: 196,665
Mix/Max/Avg. size: 1/1,471/≈ 9

---

\(^10\) Resolution limit in community detection. PNAS 2007
Facebook community structure: visualization
Outline

1. Introduction
   - Strength of weak ties theory

2. Sampling and Analysis from Facebook
   - Sampling and building a Facebook dataset
   - Facebook network analysis

3. Strength of weak ties theory evaluation
   - Nature of ties in Facebook

4. Conclusion
   - Conclusions, open problems and future work
### Strength of weak ties - distribution

**Strong & Weak**  Both distributions fit well to power laws $f(x) \propto x^{-\gamma}$ – with a decay the tail, possibly due to the sample size.

**Offset**  There exists approximately an order of magnitude of difference between number of weak and strong ties per node. Which means: every 10 links, 9 are weak and 1 is strong.

**Rich club**  This phenomenon is well-known in sociology and is called rich club. A proportion of 90%-10%, a variant of the famous 80-20 Pareto principle, often emerges in Social Networks.

![Graph showing distributions of strong and weak ties](image-url)
Strength of weak ties - tipping point

Remark According to previous experiments, the self-organization principle emerges: users aggregate in small communities highly interconnected among each other!

Tipping-point By plotting the CCDF of the probability distribution, it is possible to identify in $k \approx 5$ the tipping point from which the presence of weak ties quickly overcomes that of strong ties, making the latters less numerous in nodes with degree higher than $k$. 

![Graph showing CCDF of weak and strong ties (k)]
Interlink The inter-community links (i.e., weak ties) mainly connect communities of small size among each other. According to the **strength of weak ties theory**, weak links occur to maintain the network proficiently connected.
Ties vs Size  The probability distribution of weak ties with respect to the size of the communities follows a power law, according to the self-organization theory that explains the presence of small communities highly interconnected each other.
Strength of weak ties - discussion of results

On Facebook, most ties are weak (?)

- Only a small part of connections among individuals fall in the same community
- Do we establish friendship connections on Facebook with individuals which instead are loose friends or acquaintances?
- Do most of our friends on Facebook actually belong to different and/or far communities?

Dunbar number vs. divided attention

- Sociologists at Facebook recently assessed that on average male users actively interact with only 17 of their friends, women with 16 [Marlov] ¹¹
- Our finding tells that only 10% of connections are strong and most are weak in the Granovetter’s sense (low interaction)

Outline

1. Introduction
   - Strength of weak ties theory

2. Sampling and Analysis from Facebook
   - Sampling and building a Facebook dataset
   - Facebook network analysis

3. Strength of weak ties theory evaluation
   - Nature of ties in Facebook

4. Conclusion
   - Conclusions, open problems and future work
Conclusions, open problems and future work

Limits on Facebook:

- No access to frequency of interaction data (e.g., wall posts, chat)
- No access to the whole network, sampling hard/discouraged

Open issues in our approach:

- How to prove results are not only artifacts of community detection?
- Prove that results hold true with overlapping community detection

Future work:

- Assess role of geographical distance in tie strength
- Estimate tie strength according to an edge weight function
- Investigate symmetric vs asymmetric relations
Thanks! :-)

Questions?