MOTIVATION

- **Privacy** issues are a fundamental factor in the design, analysis, and operation of online social networks.
- Users can specify some of their contacts as private, they can form private groups, etc.
- As a result, there is not a unique social network, but instead each user has her own view of the network.

The algorithms need to respect the privacy of the users by providing results, to a given user, based only on the data that user can access.

**Naive solutions:**
1. Use only on completely public data (ineffective!);
2. Run the algorithm once for each user on a different graph (infeasible!).

~52% of NYC Facebook users hide their friends.

**THE PRIVATE-PUBLIC GRAPH MODEL**

We introduce the Private-Public Graph Model which allows to design efficient and effective algorithms for several graph problems while respecting the privacy of every user.

**Model**
- There is a public graph $G$, visible to everyone.
- For each user $u$ we have a private graph $G_u$ visible to the user.
- We want to compute a function $f(G + G_u)$ over the union of the public and private graphs for each user $u$.

**Assumption:** $G_u$ edges can be up to 2-hops from $u$ (consistent with FB, G+, etc. privacy settings).

**ALGORITHMS**

Public Graph $G$

- **Synopsis of $G$**
- Preprocessing

$$f(\ )$$

Output for $u$

Ideally, about $O(E(G))$ preprocessing time; $O(V(G))$ space and $O(E(G_u))$ query time.

**EXPERIMENTS**

<table>
<thead>
<tr>
<th>Graph</th>
<th>$A/B$</th>
<th>Cosine</th>
<th>$\tau@50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBLP</td>
<td>6.5e-3</td>
<td>99.8%</td>
<td>88.5%</td>
</tr>
<tr>
<td>LIVEJOURNAL</td>
<td>3.5e-4</td>
<td>99.1%</td>
<td>69.3%</td>
</tr>
<tr>
<td>ORKUT</td>
<td>1.6e-3</td>
<td>99.9%</td>
<td>54.6%</td>
</tr>
<tr>
<td>YOUTUBE</td>
<td>1.7e-2</td>
<td>99.8%</td>
<td>80.9%</td>
</tr>
</tbody>
</table>

Up to 4 orders of magnitude faster than the naive approach, with high accuracy.

We address several problems in this model including: Reachability, Correlation Clustering, Personalized PageRank (PPR), Affinity Scores, etc. using sampling and sketching techniques while providing provable guarantees.

Numerous interesting graph problems remain open in the private-public model.