



University of Pittsburgh

How to Measure the Information Similarity in Unilateral Relations: The Case Study of *Delicious*

Danielle Lee

June 13, 2010

International Workshop on Modeling Social Media 2010





Main Purpose

- In classic CF systems, users cannot involve in the personalization process of information.
- Users' self-defined social networks could be a good source of information and basis for personalizing information.
- There is a problem of item unit-based similarity measure. Especially, in social tagging systems, usually numeric ratings are not available (Unary or binary ratings using tags?).
- We need more semantic and multi-dimensional similarities, for social tagging systems.

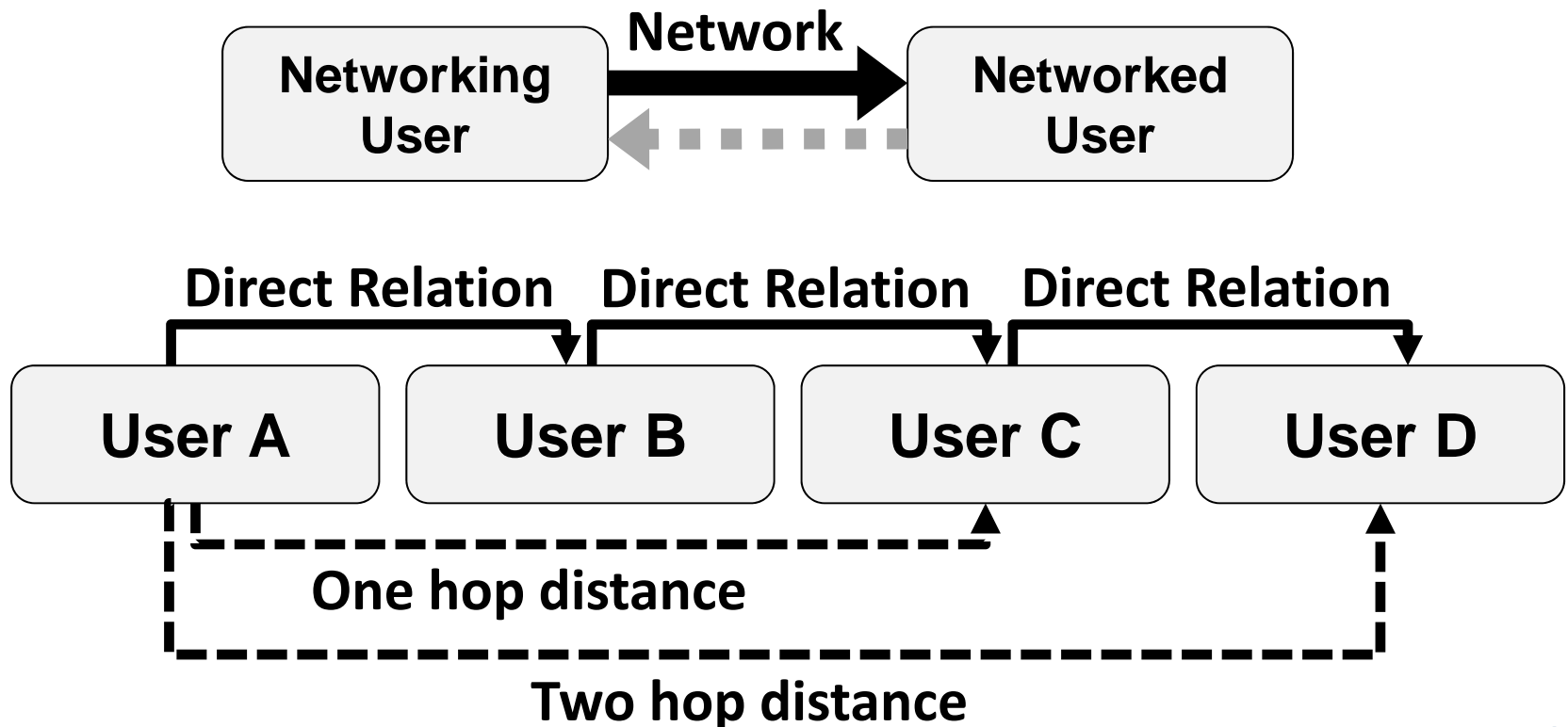


What are Unilateral Relations?

- “Following” on *Twitter*, “watching” on *CiteULike*, “contacts” on *Flickr*, or “network” on *Delicious*
- Users’ self-pointed connections are based upon their utility of information
- Does not require mutual agreement about being in the relationship
- The relationship is not always reciprocal
- One of the newly emerging “less bounded” relationships (Wellman, 2001)
- Java, et al. (2007) : Three kinds of users’ intention to socialize on the twitter – friendship, information sharing and information seeking.

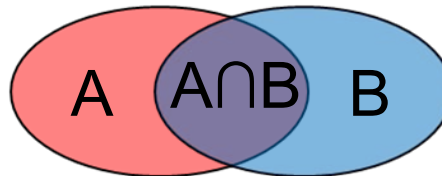
Relationships in Consideration

- Direct & Indirect Unilateral Relations



Measures of Similarity

- Item-unit, Metadata-unit, Macro & Micro Tag-unit Similarities
 - Metadata Similarity: the similarity of root address of each bookmark
 - Micro-Tag Similarity: tag similarity for a common item
 - Macro-Tag Similarity: the similarity of the overall tags regardless of the tagged items



Absolute Numbers = $A \cap B$

Inlink Power = $(A \cap B) / A$

Outlink Power = $(A \cap B) / B$

Overall/Jaccard Power = $(A \cap B) / (A \cup B)$



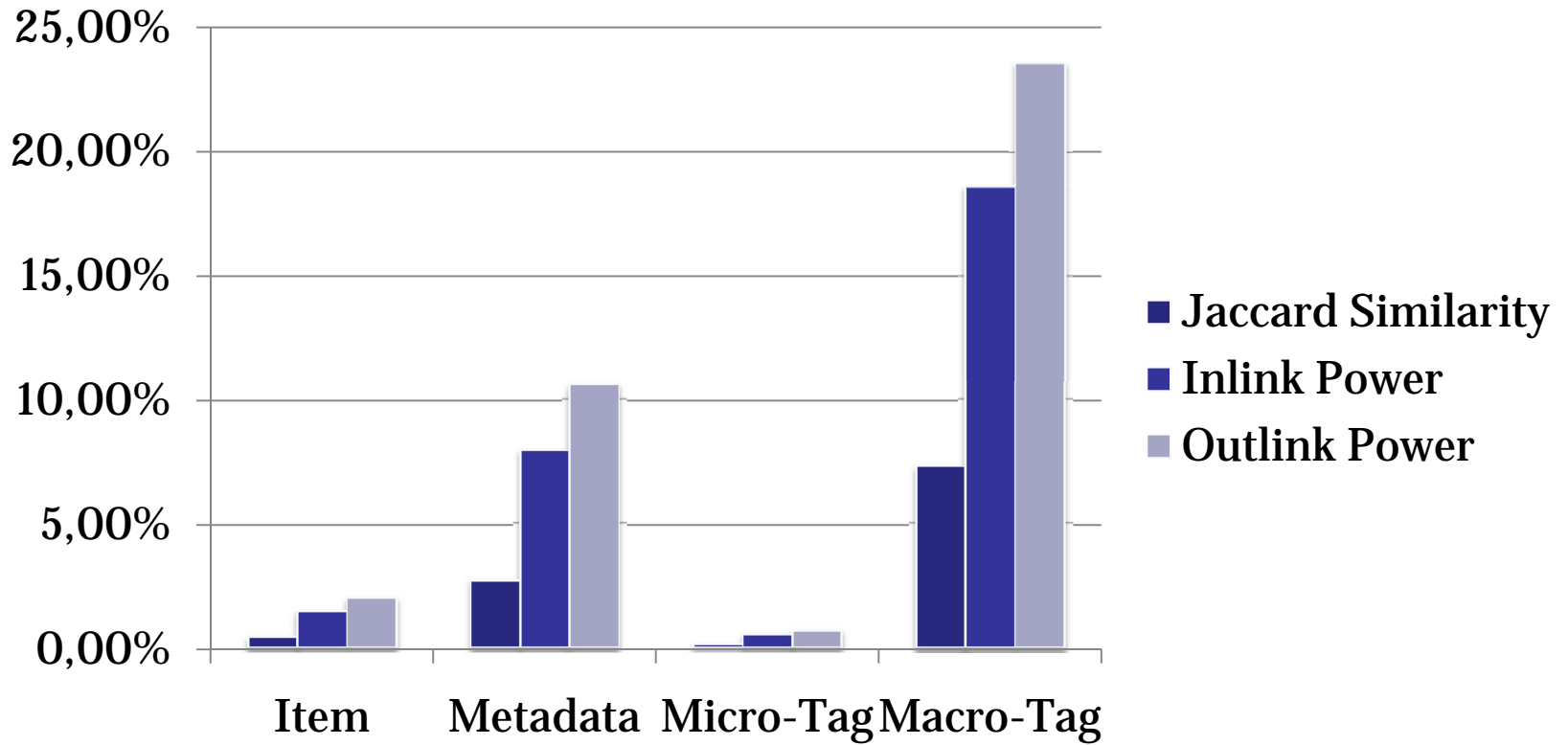
The Data Set

- Collected *Delicious* data randomly using snowball sampling.

| | |
|--|---------|
| Total No. of Users | 11772 |
| Total No. of Distinct Items (Bookmarks) | 5191538 |
| Avg. No. of Items per User | 788.83 |
| Total No. of Distinct Tags | 651622 |
| Avg. No. of Tags per User | 2268.87 |
| Total No. of Unilateral Relations | 16538 |

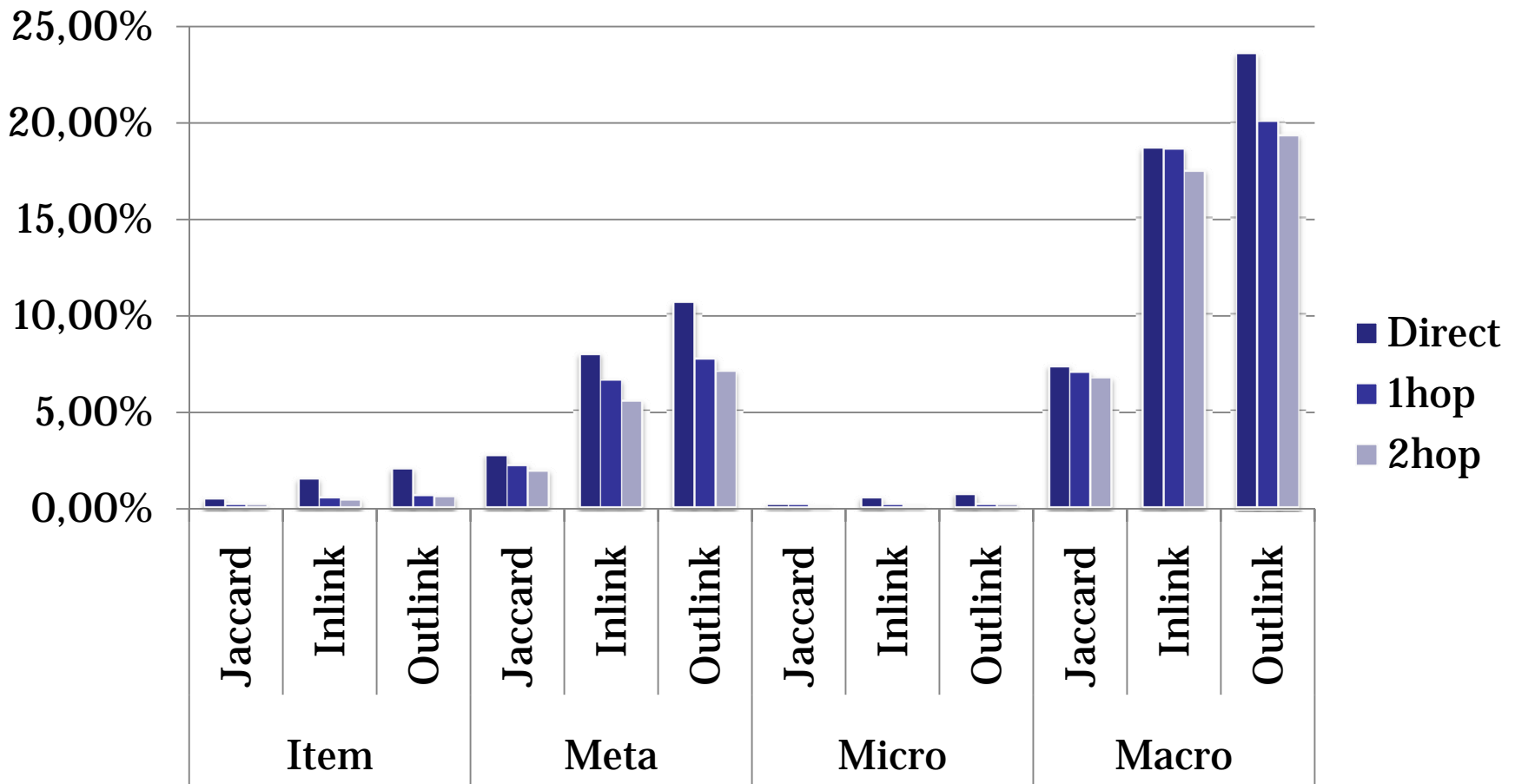


Results of Direct Relations





Result of Relations with Distances





Conclusion

- In *Delicious* social tagging system, users undershare items
- However, they shared more semantically rich information such macro-tags and metadata.
- The information sharing pattern was consistently shown in the results for the indirect relations

