Event-based Social Networks: Linking the *Online* and *Offline* Social Worlds

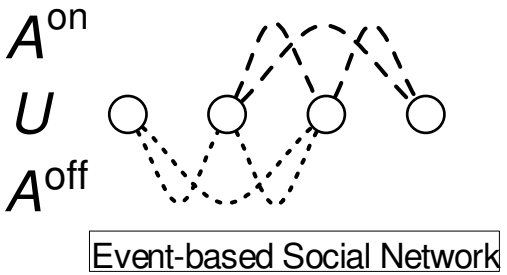
Xingjie Liu[*], Qi He[#], **Yuanyuan Tian**[#],
Wang-Chien Lee[*], John McPherson[#], Jiawei Han[+]
The Pensalvania State University[*],
IBM Almaden Research Center[#],
University of Illinois at Urbana-Champaign[+]

Event-based Social Services

- Examples: 


PLANCAST
  - Create social events, join events through RSVP
  - Communicate with each other online (like in Facebook and Twitter)
- Linking the ***online*** and ***offline*** social worlds
  - ***Online*** virtual world: exchange thoughts and share experiences
  - ***Offline*** physical world: face-to-face social interactions in events
    - ***when*** and ***where***, ***who*** and ***who*** did ***what*** together
- Advantages over conventional social networks
  - Stronger social ties
    - Participating in a physical activity together >> befriends online
  - Stronger intents
    - Participating in a ski event >> talking about skiing online

Event-base Social Network (EBSN)



Event-based Social Network

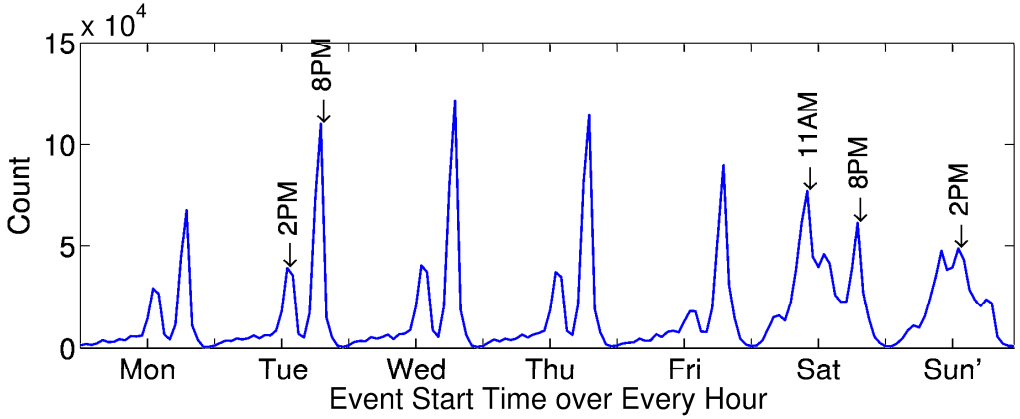
- EBSN definition: G=<U, A^{on}, A^{off}>
  - U: users
  - A^{on}: ***online*** social interactions
  - A^{off}: ***offline*** social interactions
  - Additional attributes can be associated with users and interactions
  - G^{on}=<U, A^{on}> forms an ***online*** social network
  - G^{off}=<U, A^{off}> forms an ***offline*** social network
- Comparison to Location-Based Social Network (LBSN), e.g. Foursquare and Gowalla
  - Similar online virtual world
  - But offline part of LBSN only records ***individual*** place checkin behaviors
- Crawled Datasets
  - **Meetup** (EBSN): 5.15 million users, 5.18 million events, 42.7 million RSVPs, 97.6 thousand groups, 10.7 million memberships
  - **Gowalla** (LBSN): 566 thousand users, 36.8 million checkins, 2.84 million locations, 2.43 million links

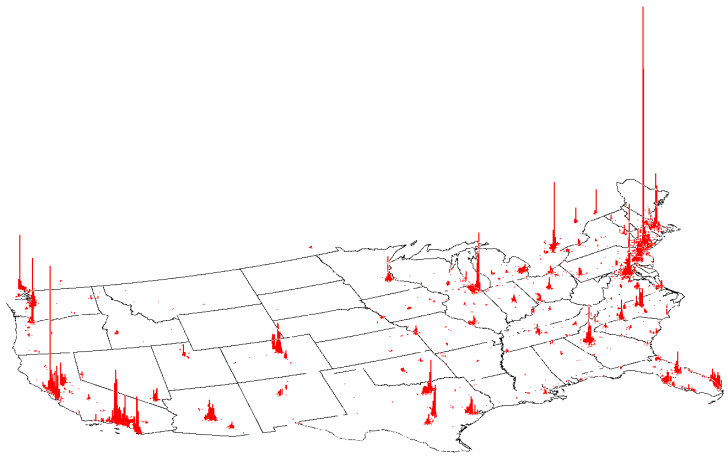
Questions about EBSNs

- What are the unique features of an EBSN?
  - Online + offline
  - Are they correlated?
- How to detect the communities in an EBSN?
  - Online and offline interactions together define communities
  - Are the communities more cohesive than in other SNs?
- How does information flow inside an EBSN?
  - Information can flow online and offline
  - e.g. event recommendation
    - Event has very short life time – cold start problem

Properties of Meetup EBSN (1/2)

- Events show regular temporal and spatial patterns

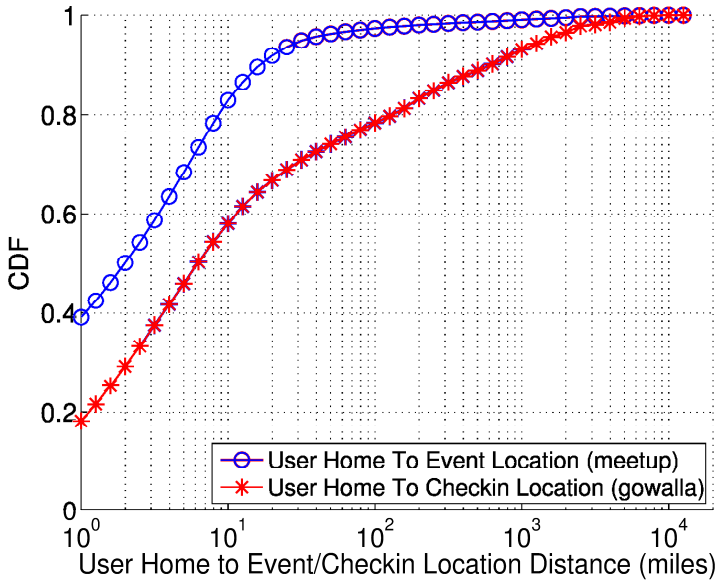


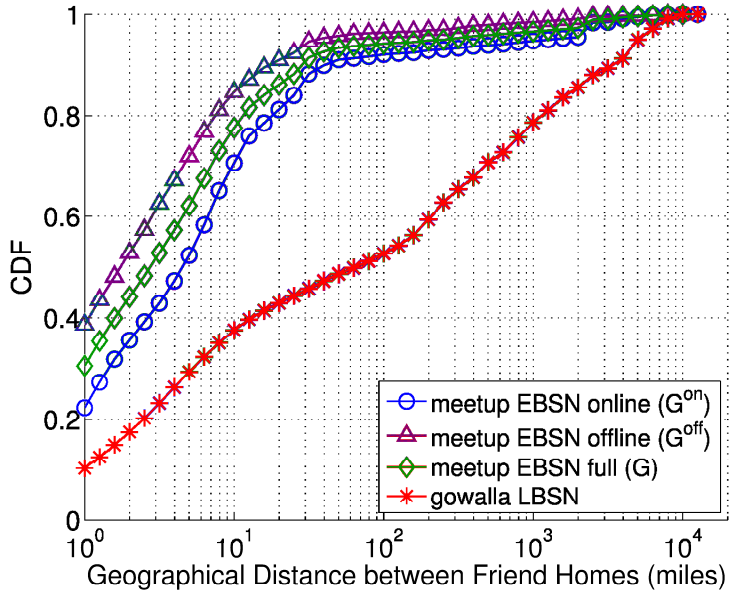


- Correlation between online and offline networks
  - Degree correlations: 0.37, cluster coefficient correlation: 0.39

Properties of Meetup EBSN (2/2)

- Strong locality of social events and interactions
  - 81.93% events of a user are within 10 miles of his/her home* location
  - 84.61% of offline friends live within 10 miles to each other





* By discretizing the world into 25km X 25km cells, a user's home location is identified as the averages position in the most frequent cell.

Community Detection in EBSNs

- Communities are defined by both online and offline interactions.
- Approach 1: Linear Combination
  - $A = \gamma * A^{\text{on}} + (1-\gamma) * A^{\text{off}}$, then apply Fiedler method
- Approach 2: Generalized SVD
  - $A^{\text{on}} = \mu \Sigma_1 Y^T, \;\; A^{\text{off}} = Y \Sigma_2 \nu^T$
  - 2nd to m-th smallest singular vectors in Y as a m-1 dimensional vector for a K-means
- Approach 3: Extended Fiedler Method
  - Objective function: (extension of normalized cut)

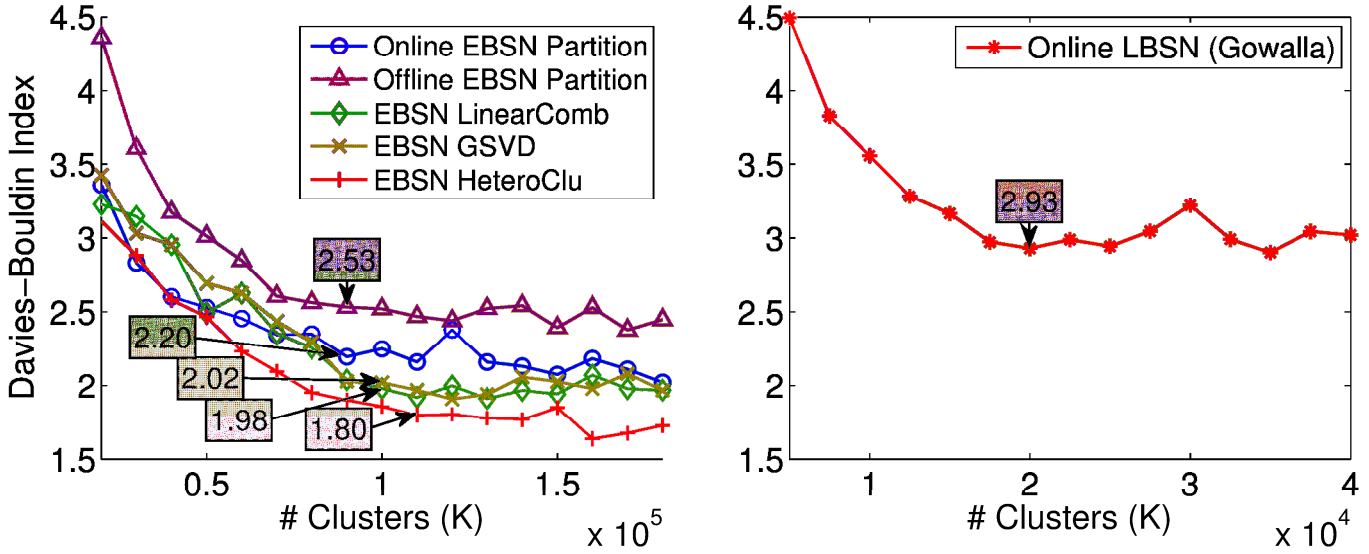
$$\min \; \alpha \frac{\mathbf{y}^T (D^{\text{on}} - A^{\text{on}}) \mathbf{y}}{\mathbf{y}^T D^{\text{on}} \mathbf{y}} + (1-\alpha) \frac{\mathbf{y}^T (D^{\text{off}} - A^{\text{off}}) \mathbf{y}}{\mathbf{y}^T D^{\text{off}} \mathbf{y}}$$

$$subject\ to\ \mathbf{y}^T D^{\text{on}} \mathbf{1} = 0, \mathbf{y}^T D^{\text{off}} \mathbf{1} = 0, \mathbf{y} \neq 0$$

  - A heuristic algorithm (**HeteroClu**)
    - A bottom-up loose clustering based on linear combination
    - A top-down recursive binary-cut

Cohesiveness of Communities

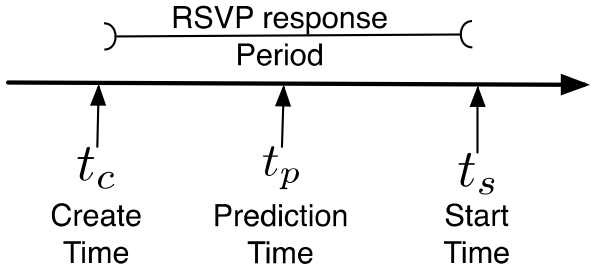
- Cohesiveness: how similar the users in a community are compared to other communities
  - Utilize user tags (specify user interests)
  - Measure: Davies-Bouldin index (smaller value is better)



- **HeteroClu generates much better communities.**
- **Communities in Meetup EBSN are more cohesive than in Gowalla LBSN.**

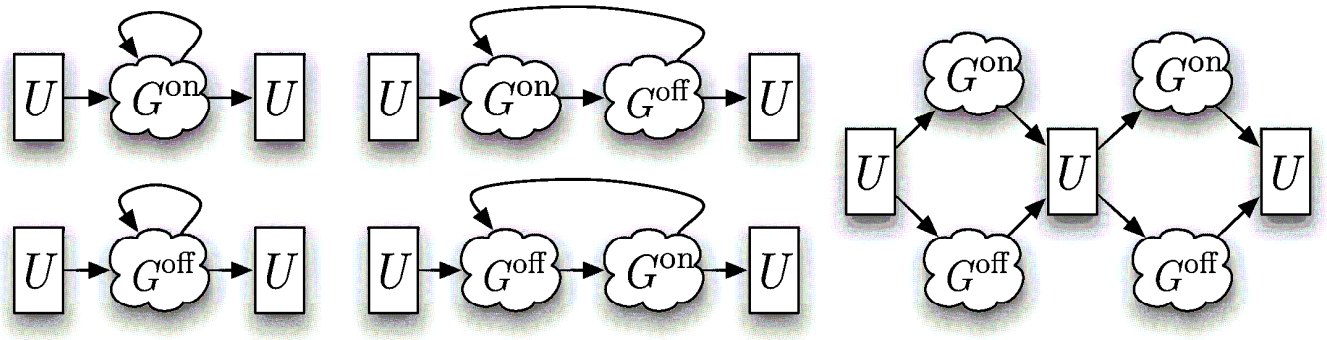
Information Flow in EBSNs

- Application: Event Recommendation
  - For a given event, recommend users to participate
  - Event has very short life time
    - Given creation time t_c, start time t_s, recommendation is only valid for $t_c<t_p<t_s$.
      - RSVPs between t_c and t_p, are training examples.
    - Existing supervised learning algorithms perform poorly due to the cold start problem.
  - Study event recommendation using information diffusion approaches



Information Flow Patterns

- Typical information diffusion in an EBSN



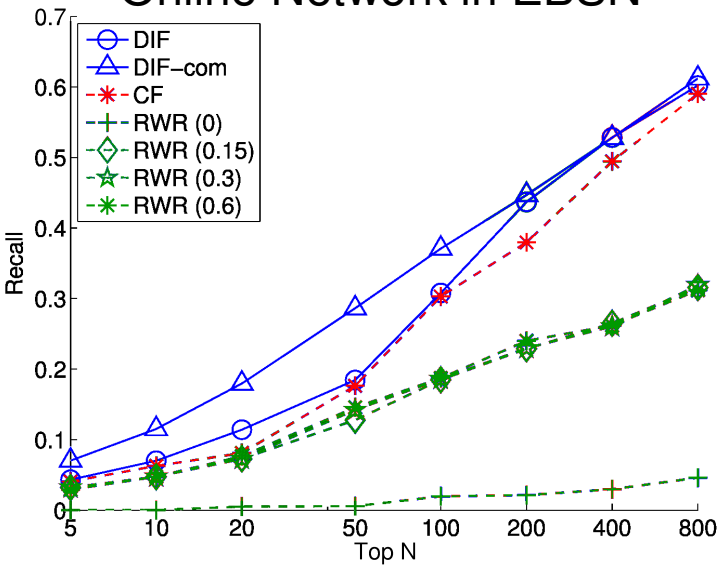
(1) single channel (2) cascaded channels (3) paralleled channels

- Community-based diffusion
  - Information tends to, but not restricted to flow in its own community

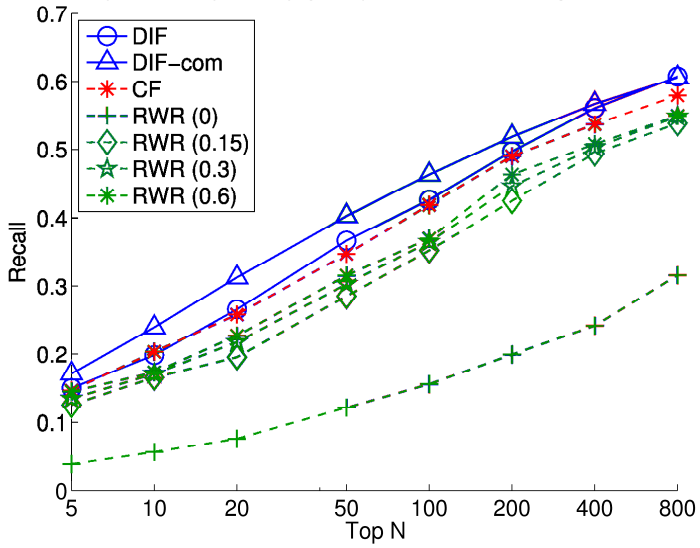
Event Recommendation on Homogenous Networks

- Baselines:
  - CF: collaborative filtering
  - RWR: random walk with restart for different restart probabilities
- Divide data 8:2 and recommend valid events

Online Network in EBSN

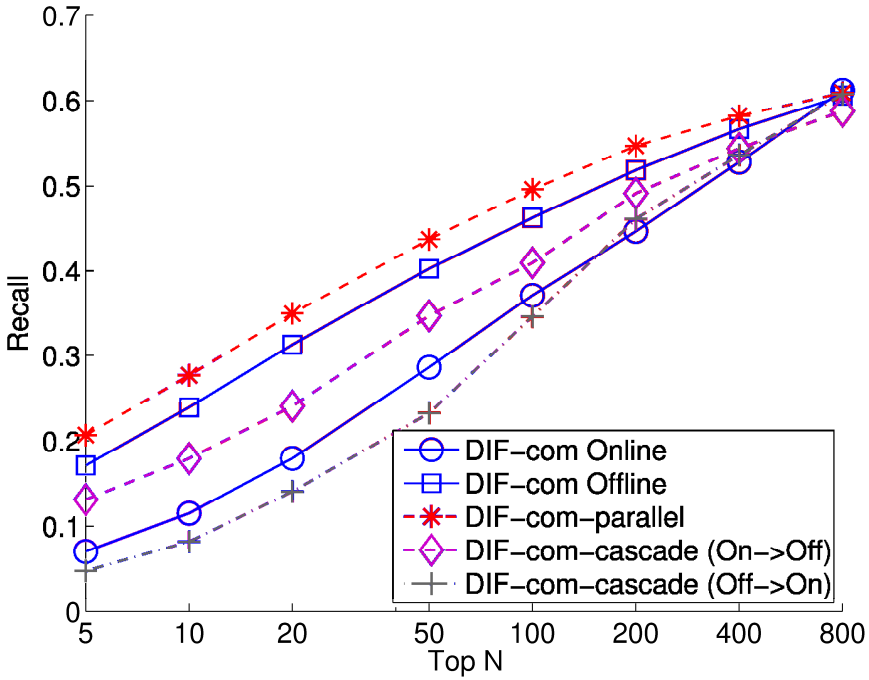


Offline Network in EBSN



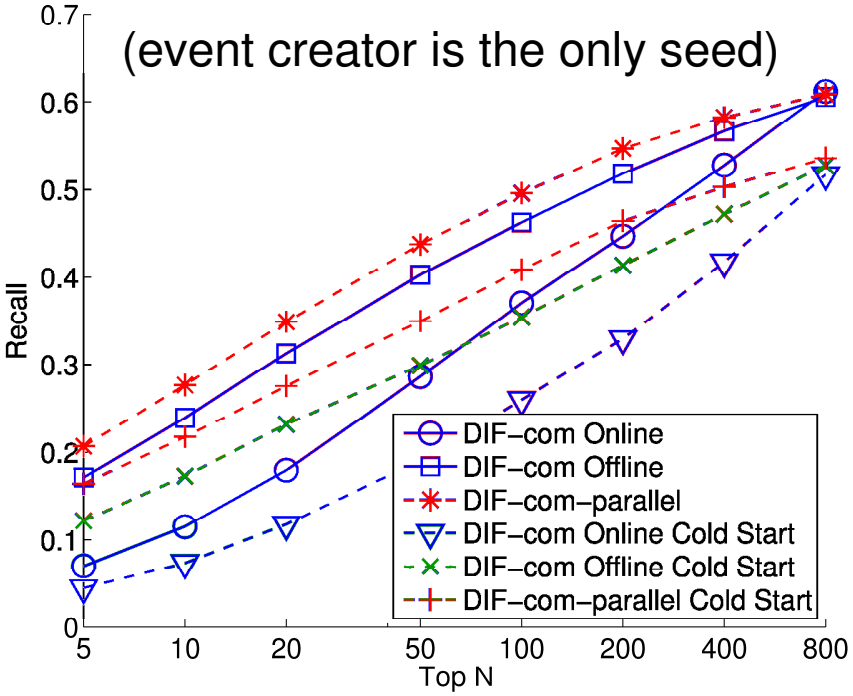
- **Community-based diffusion works better than CF or RWR for event recommendation**

Diffusion on EBSNs



Effect of Cold Start

(event creator is the only seed)



- **Community-based parallel diffusion model produces the best recall**
  - **Even satisfactory with extreme cold start.**

Future Work

- More EBSN examples
- Generative model for EBSNs
- Better clustering algorithms on EBSNs
- More applications on EBSNs

- Special Thanks
  - Prof. Jon Kleinberg for providing us a comprehensive list of related works

- Datasets to download
  - http://www.largenetwork.org/ebsn

Thank You

- Questions?
- Suggestions?