Maximal Labelled-Clique and Click-Biclique Problems for Networked Community Detection

Debajyoti Bera
IIIT-Delhi, India
INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI

Flavio Esposito
Meghan Pendyala
Saint Louis Univ., USA
SAINT LOUIS UNIVERSITY
EST. 1818

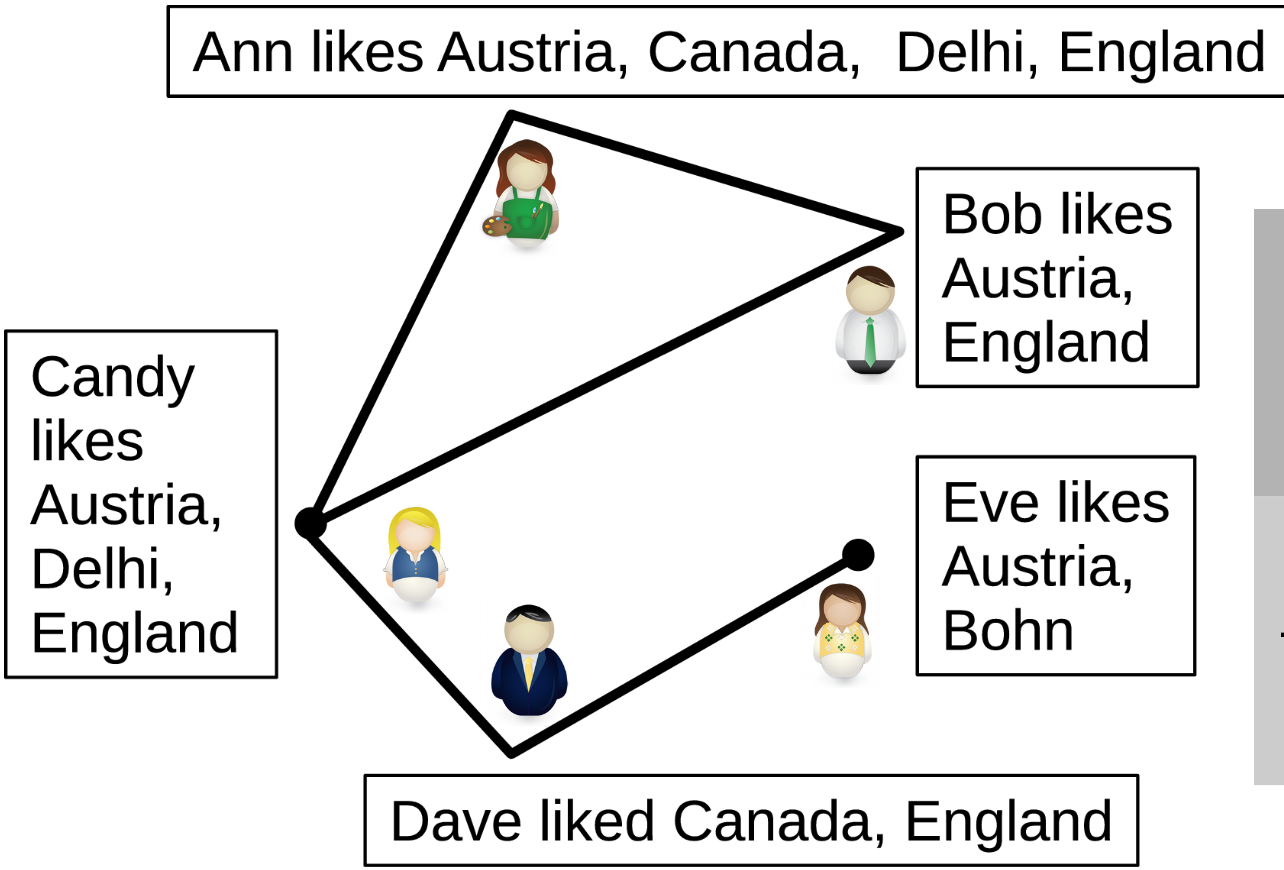
Globecom 2018, SAC-SN
10th December 2018

Social Network + User Interests

+

Q: Find better communities in networks with discrete attributes using graph-theoretic modelling

Like-Minded Communities



Ann and Candy are friends to each other and both like Austria, Delhi, England

Ann, Bob and Candy are friends to each other and all like Austria, England

A recently proposed technique

Given a social network along with interersts associated with each user, find communities:

- Group of users that are ...
- Like-minded (have a common set of "likes")
- Highly-connected (connected to each other)

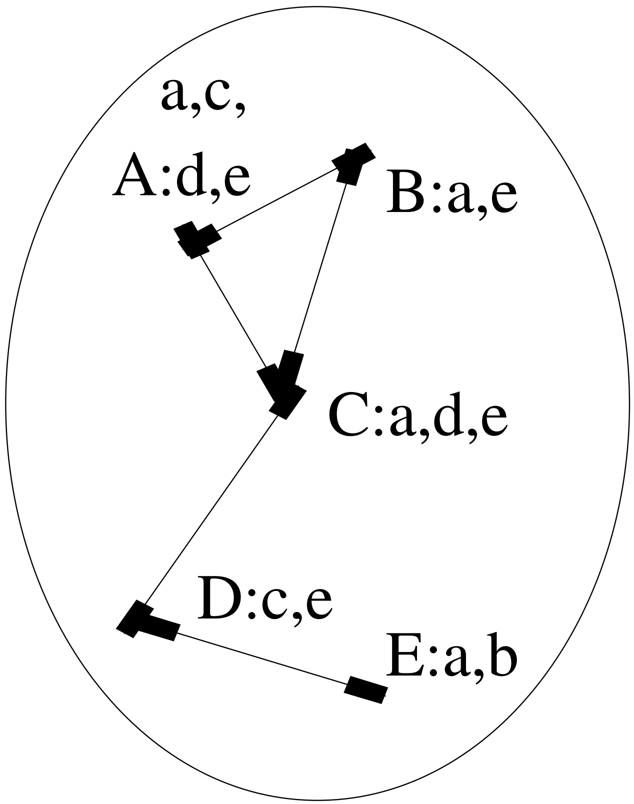
Can be found using maximal-clique enumeration and frequent-itemset mining (multiple times).

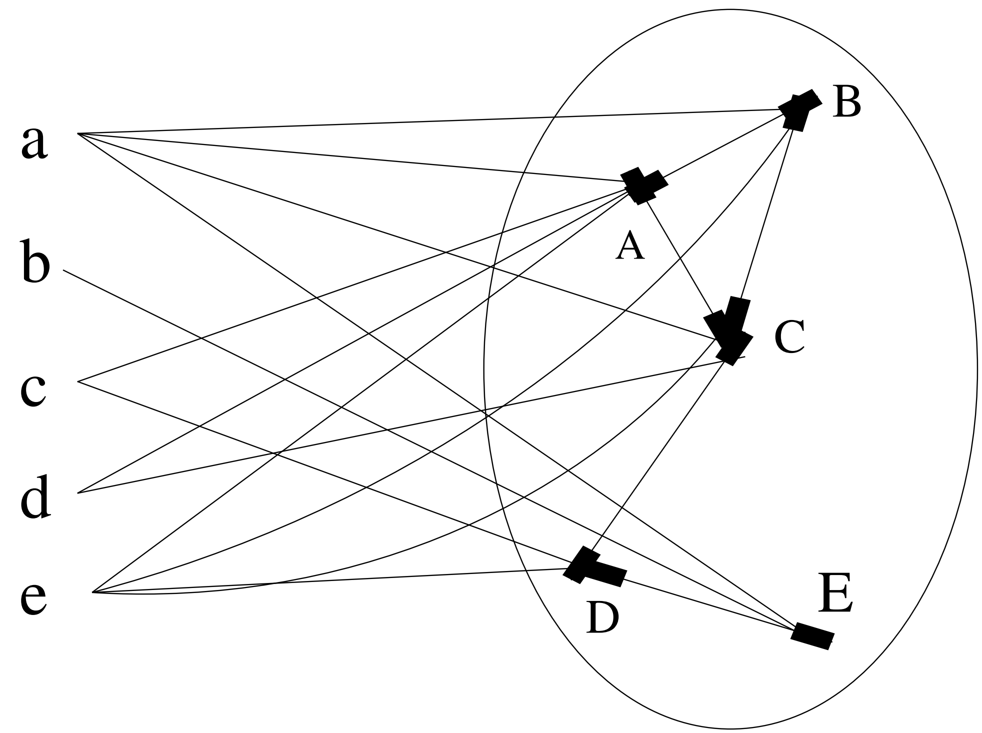
[Like-minded communities: bringing the familarity and similarity together, Modani et al., World Wide Web journal, 2014]

Our Results : Single graph algo.

- Define graph-theoretic pattern "Click-Biclique" (CBC)
  - Use "maximal" CBCs to model like-minded highly-connected communities in networks with labels
  - Relate maximal CBC mining to network clustering in various types of networks
- Give two graph-theoretic algorithms for listing all maximal CBCs
- Emperically evaluate these algorithms on real & synthetic datasets

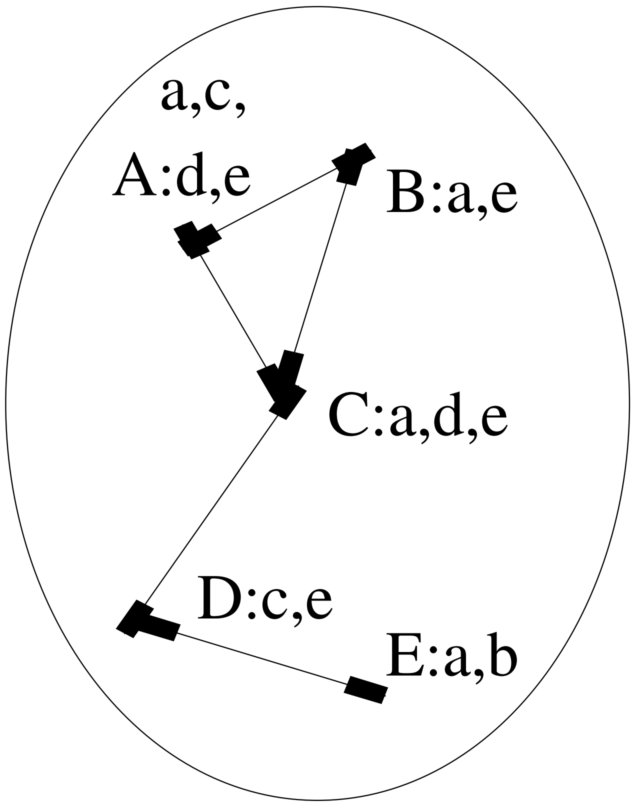
Labelled Graph

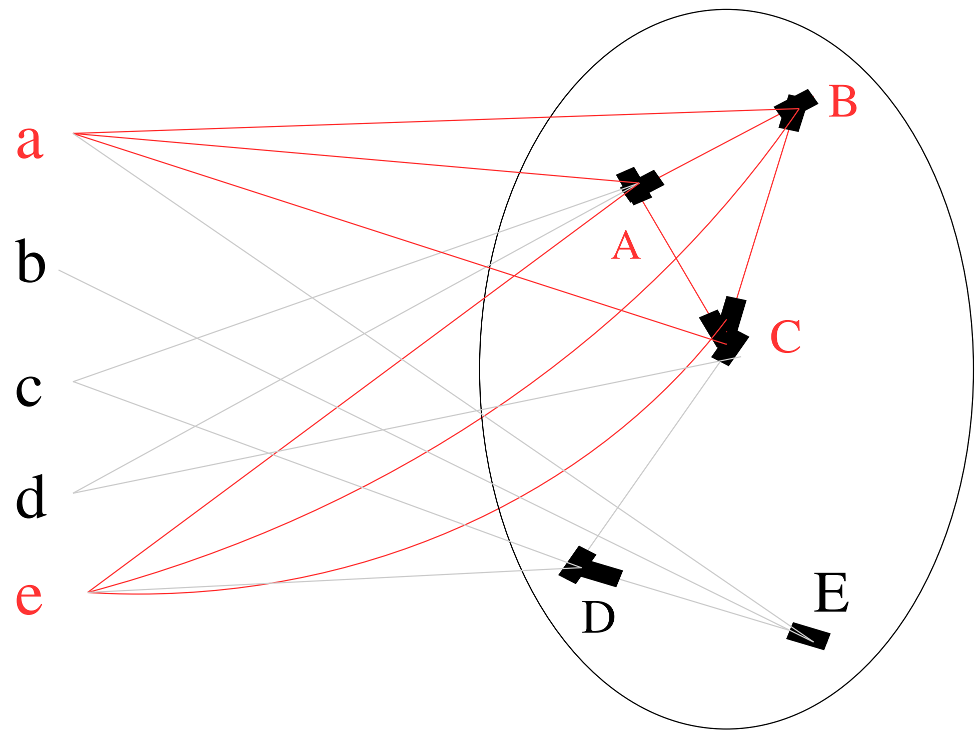




Labelled-clique: Subset of labels L and subset of vertices V such that V forms a clique and every v in V has every label in L.
Maximal labelled maximal clique: <L,V> *s.t.* both L and V cannot be increased.

Labelled Graph

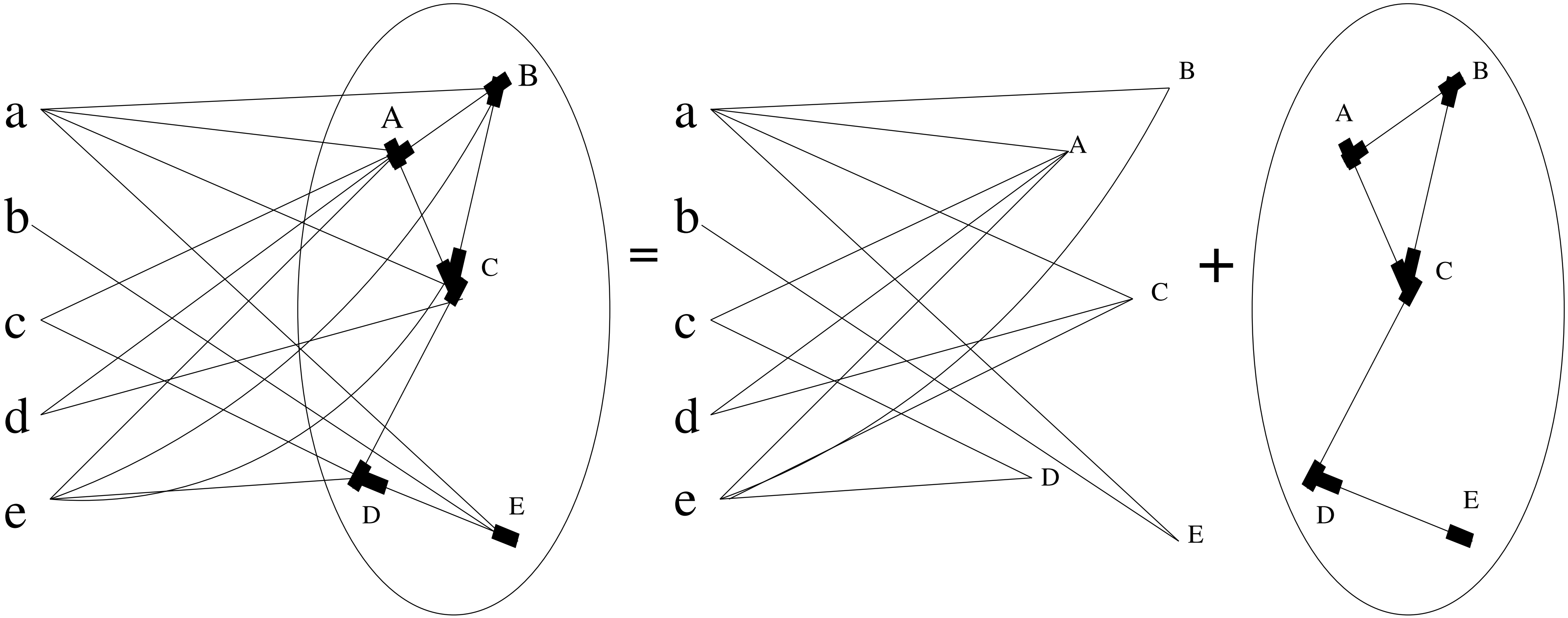




<u>Labelled-clique</u>: Subset of labels L and subset of vertices V such that V forms a clique and every v in V has every label in L.
<u>Maximal labelled maximal clique</u>: <L,V> *s.t.* both L and V cannot be increased.

Example: < L = {a,e}, V = {A,B,C} >   < L = {a,d,e}, V = {A,C} >

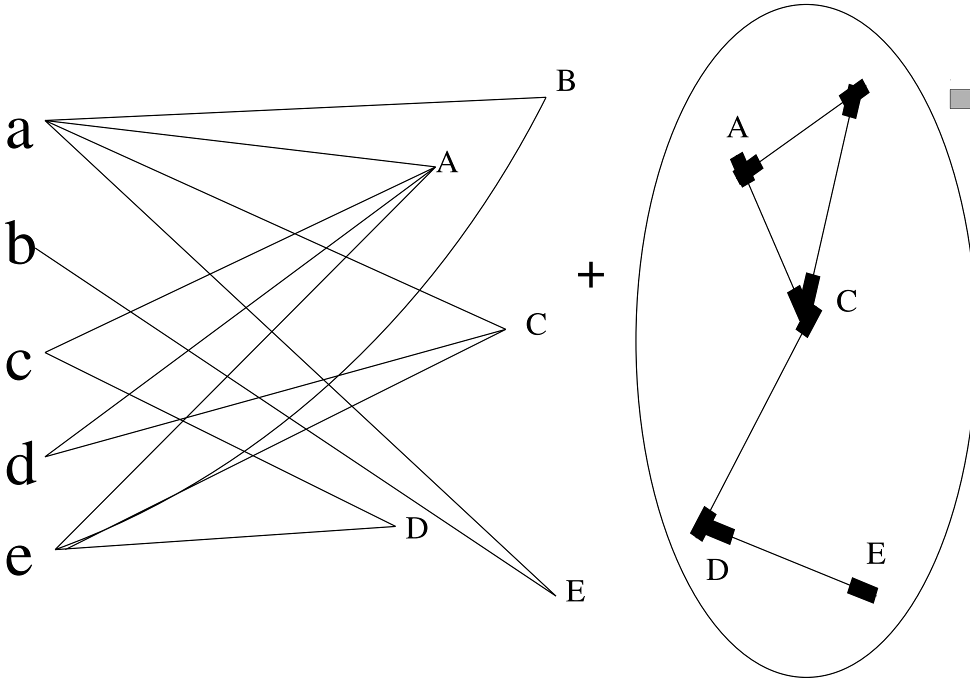
Joined Graph



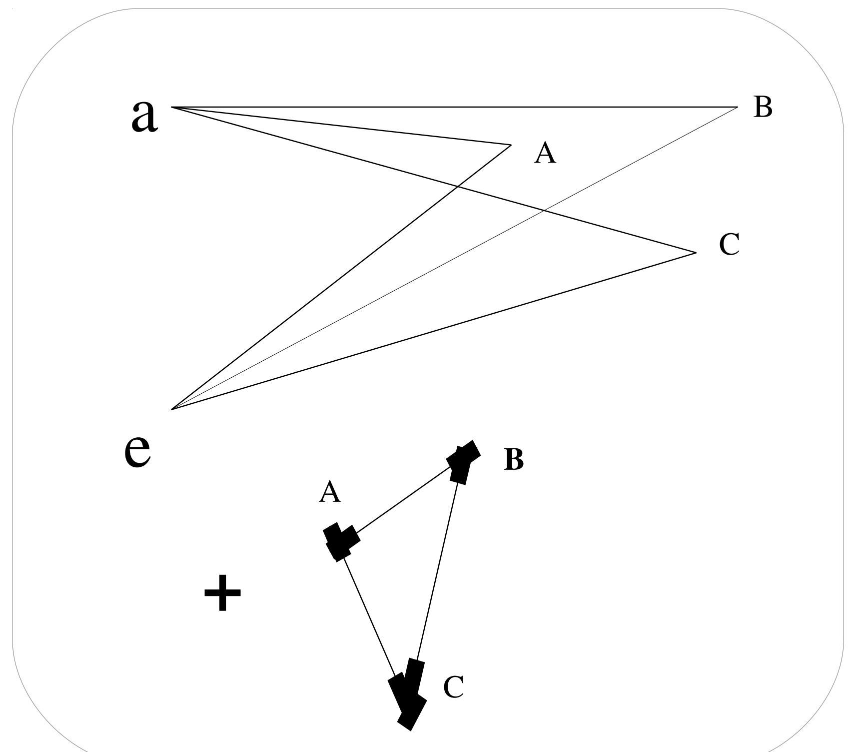
Labelled-graphs can be modelled as joined-graphs.

Debajyoti Bera, IIIT-Delhi
New Delhi, India

Maximal CBC



Maximal-labeled maximal-clique:
L = {a,e}, V = {A,B,C}



Clique-Biclique (CBC):
- Biclique {a,e} x {A,B,C} +
- Clique {A,B,C}

Maximal-labeled maximal-clique ≡ Maximal clique-biclique

CBCs for network analysis

- Targetted advertising: groups of users who are <u>good friends of each other</u> **and** <u>share a lot of interests</u>

- Cyber-physical systems and fog-computing: for task allocation to <u>sensors that are well-connected to each other</u> **and** <u>possess similar computing abilities (battery capacity, camera, computing libraries, etc.)</u>

- Gene network: finding sets of <u>genes that influence each other</u> **and** <u>have a common set of phenotypes</u>

Enumeration of maximal CBCs

There can be too many CBCs, hence, we find simply want to find maximal ones.

Related enumeration algorithms

- Maximal clique enumeration: variants of Bron-Kerbosch (BK) runs well in practice, asymptotically better algorithms also exist
- Maximal biclique enumeration: a few algorithms have been proposed, state of the art is iMBEA

[BK] Finding all cliques of an undirected graph, Bron & Kerbosch, CACM, 1973

[iMBEA] On finding bicliques in bipartite graphs: a novel algorithm ..., Zhang et al., BMC Bioinformatics, 2014

Enumeration algorithms

Maximal CBCs cannot be obtained by post -processing maximal cliques or maximal bicliques. Approach by Modani et al. involves running multiple algorithms several times with pre & post processing.

- mCx : Run a maximal clique enumeration algorithm on a modification of a joined-graph
  - Inefficient since the modification can be expensive and may lead to generation of redundant maximal clique
- mCBC : Backtracking-search like recursive algorithm along the same lines of BK and iMBEA

Emperical evaluation on synthetic datasets to understand how mCx and mCBC perform with respect to labellings.

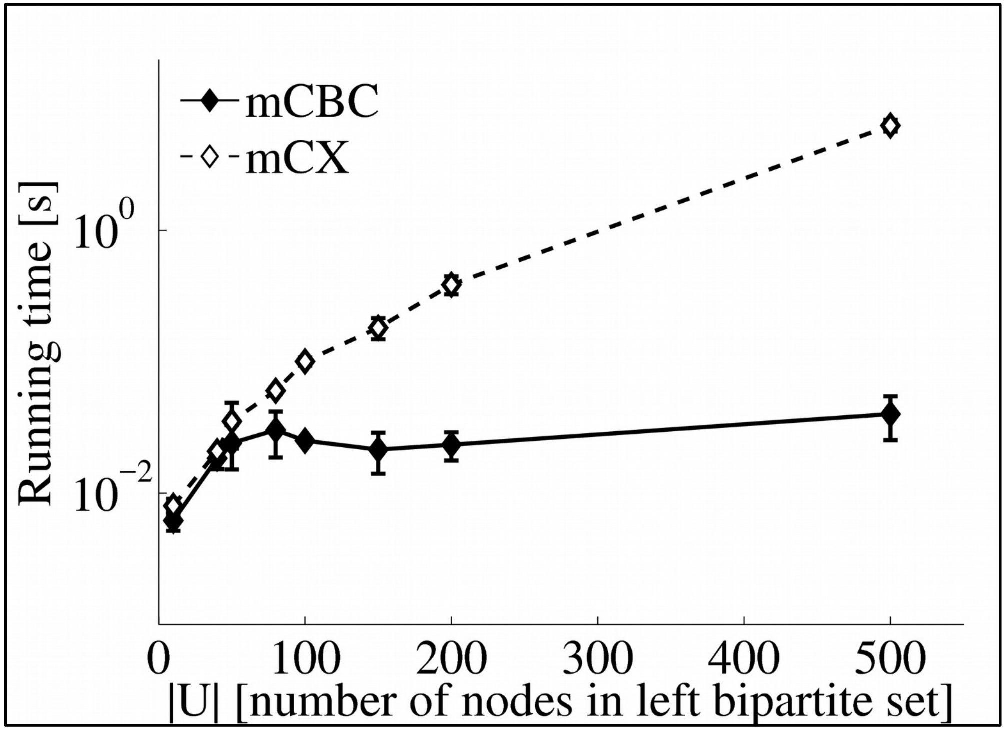
- Number of labels
- Density of labellings

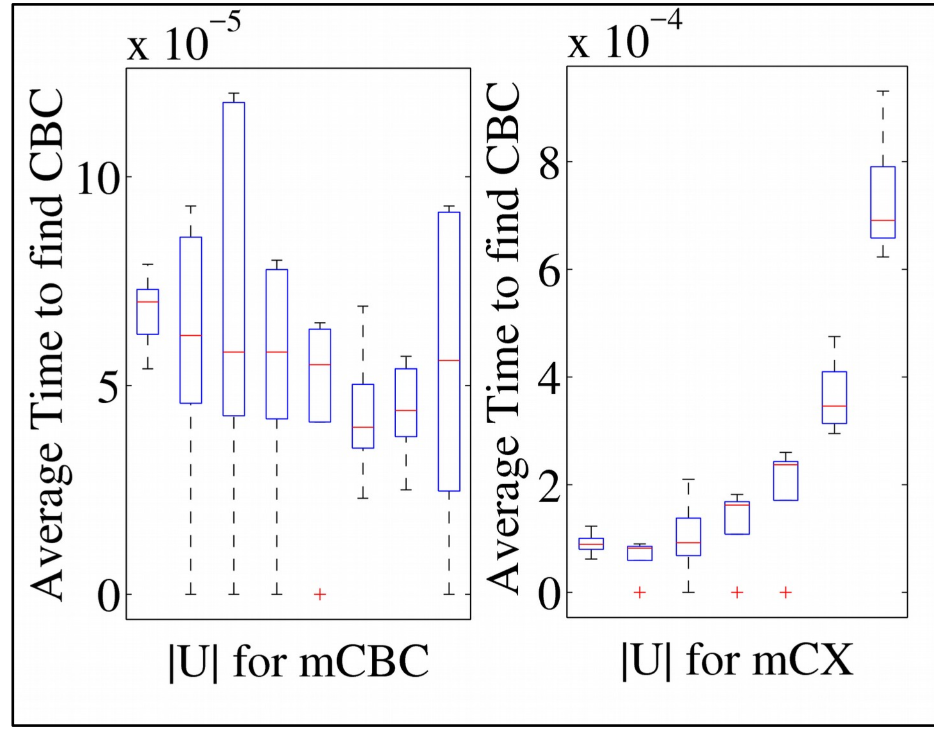
(Using Erdos-Renyi random graphs)

I. With more labels, mCBC becomes even better.

II. With more avg. labellings per user, mCBC is better but not by much. (Happens since any maximal clique on users will most likely induce a maximal biclique, and hence, a maximal CBC.)

Interesting Observation





Performance of mCBC remains unaffected by increase in the number of labels whereas the performance appears to suffer in an <u>exponential</u> manner for mCx.

Real social-network datasets

Property	Café TheMarker's	Ning Creator's
Number of users	6,333	11,011
Number of groups	88	81
Number of inter-user links	19,315	76,263
Number of clique	21184	42,320
Number of bicliques	553558	2,002
Number of clique-bicliques	19,917	5,459
Users with unique "taste": Number of maximal CBCs with 1 user and 1 label	2,285	1,049
Popular groups: Number of maximal CBCs with at least 3 users and at least 3 labels	2,211	118

(Maximal) CBC analysis can be used to characterize different (social) networks.

Since then …

- Improved version of mCBC.
- Better usage of maximal CBCs in finding like-minded communities compared to that of Modani et al.
- Developed quantifiable metrics to characterize social network based on their inter-user links and labelings.
- Relate Maximal CBC enumeration to the problem of mining frequent patterns.

That's All!
Thanks.