Complex Data Mining for
Human Interaction Analysis
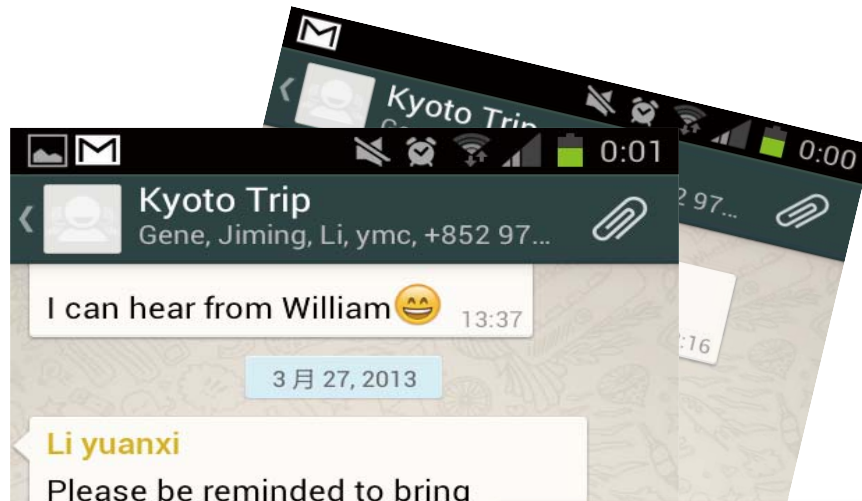
William K. Cheung (張國威)

Department of Computer Science

Hong Kong Baptist University

Digital Traces ...

Digital Traces …

- Digital Traces *Left By* Life
  - Ease data collection for *studies on human behavior.*

- Digital Traces *Shaping* Life
  - *Shape* new behaviors (better or not?)
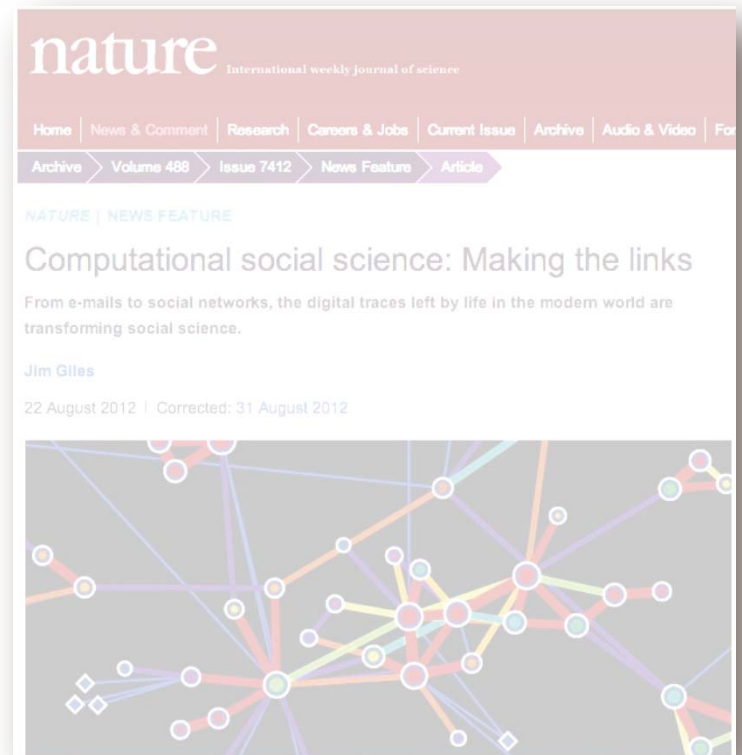
Complex Data ....

www.orgnet.com

Octopus Card
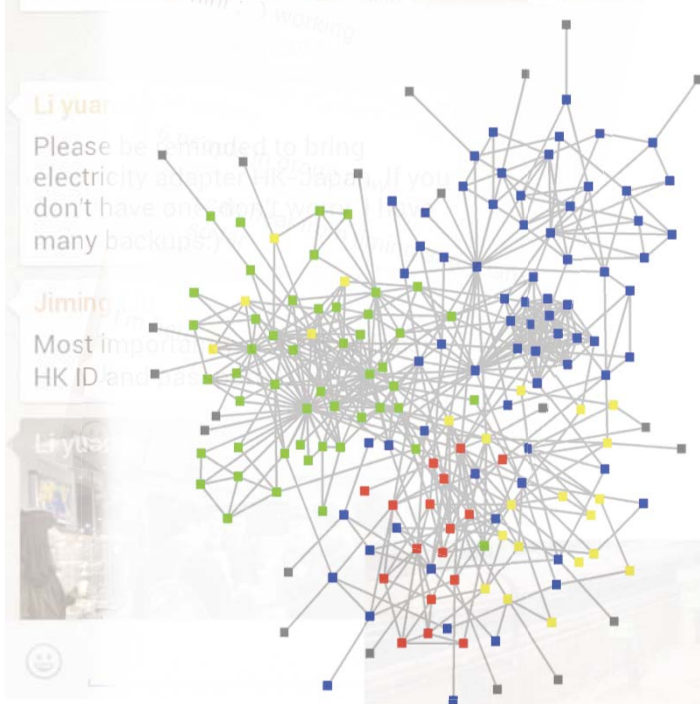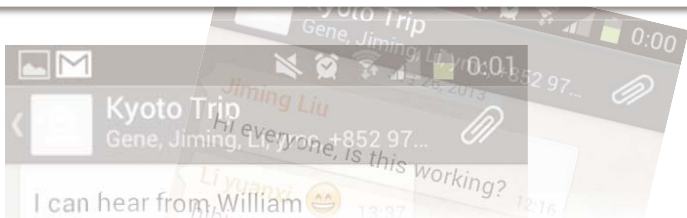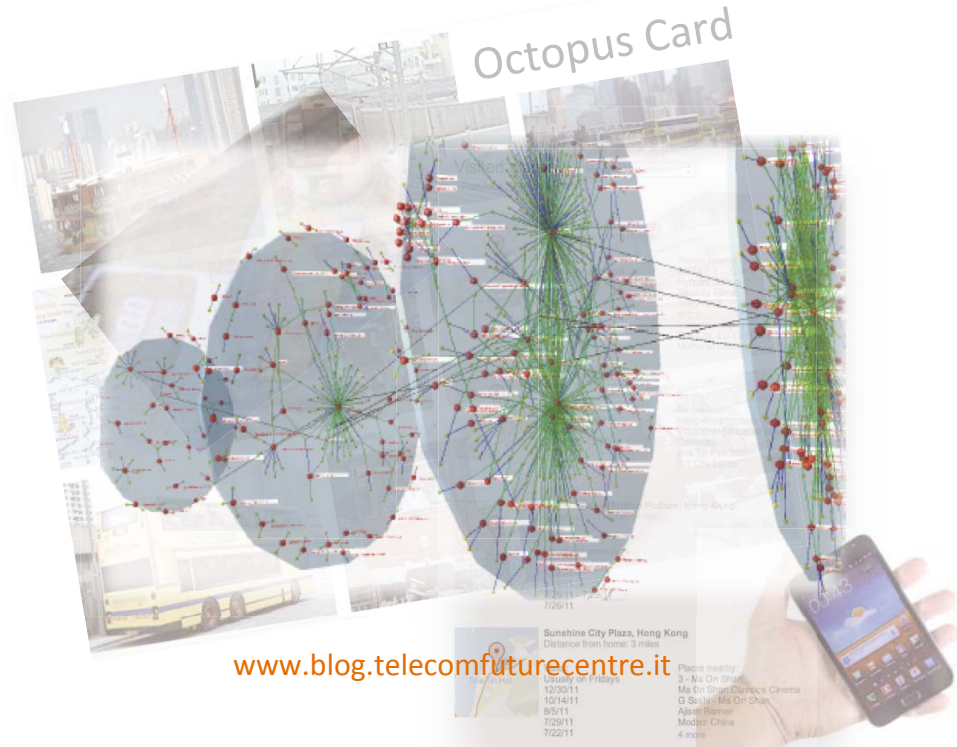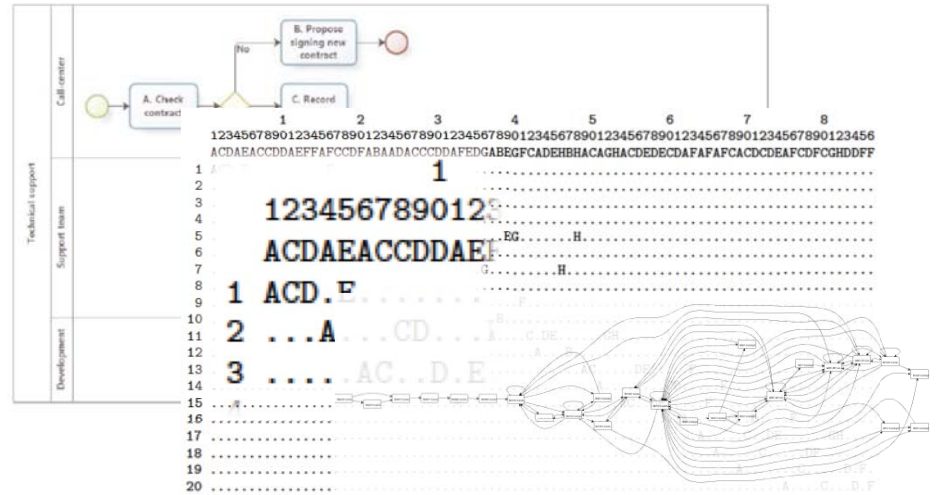
www.blog.telecomfuturecentre.it

Focus 1 - Human Communication Mechanisms
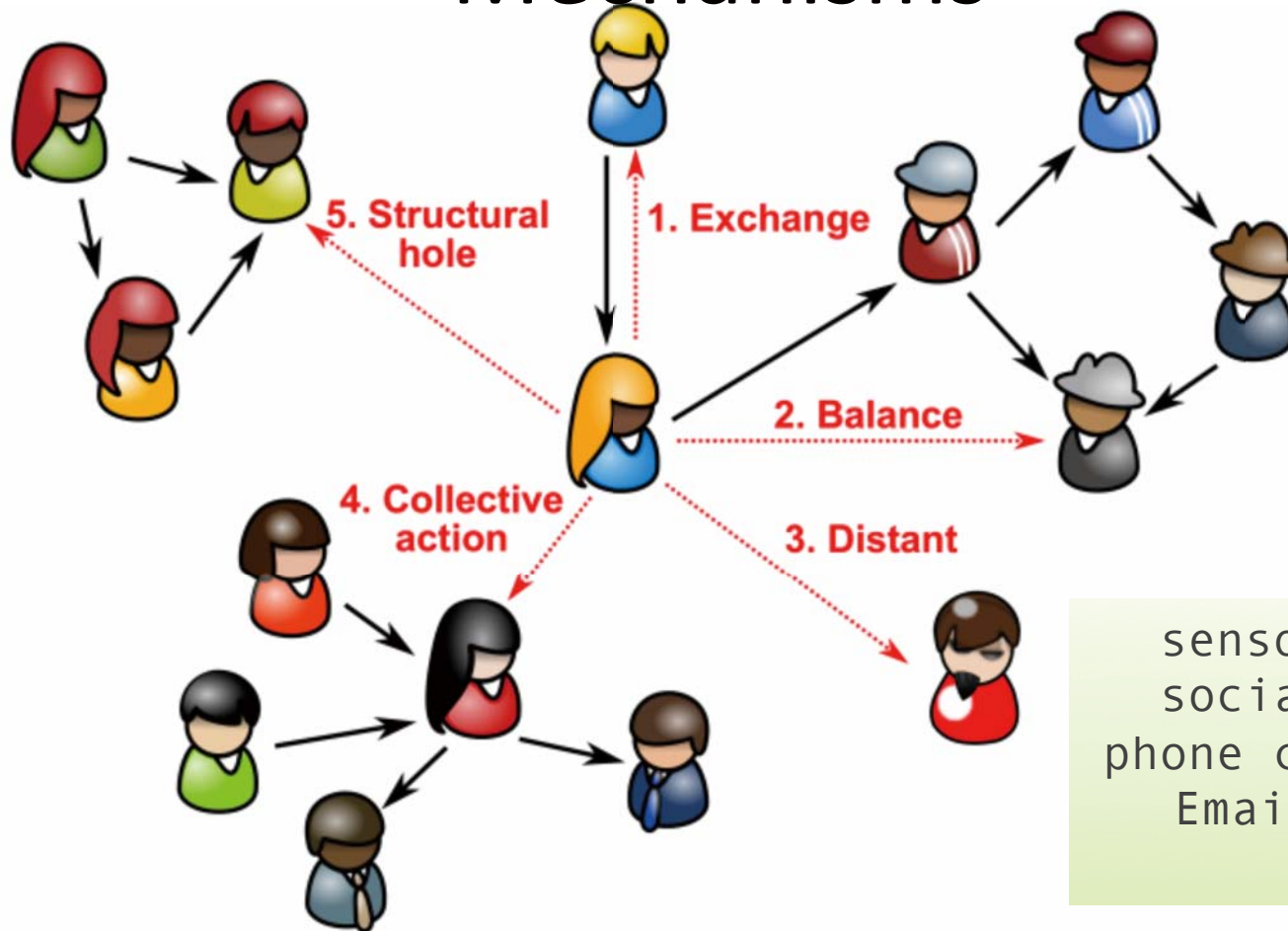
sensor network
social network
phone call network
Email network
...

L. K. Gallos, S. Havlin, F. Liljeros, H. A. Makse, "How people interact in evolving online affiliation networks", Phys. Rev. X 2, 031014 (2012)

P1 - Human Communication Mechanisms
Stochastic + Network Motifs

Finite Mixture Models

$$p(X_w|\Theta) = \sum_{m=0}^{k} \alpha_m p(X_w|\Theta_m)$$

multinomial, Markov chain, …

α_1 α_2 α_K

Over-represented Subgraph Pattern #1   Over-represented Subgraph Pattern #2   . . .   Over-represented Subgraph Pattern #K
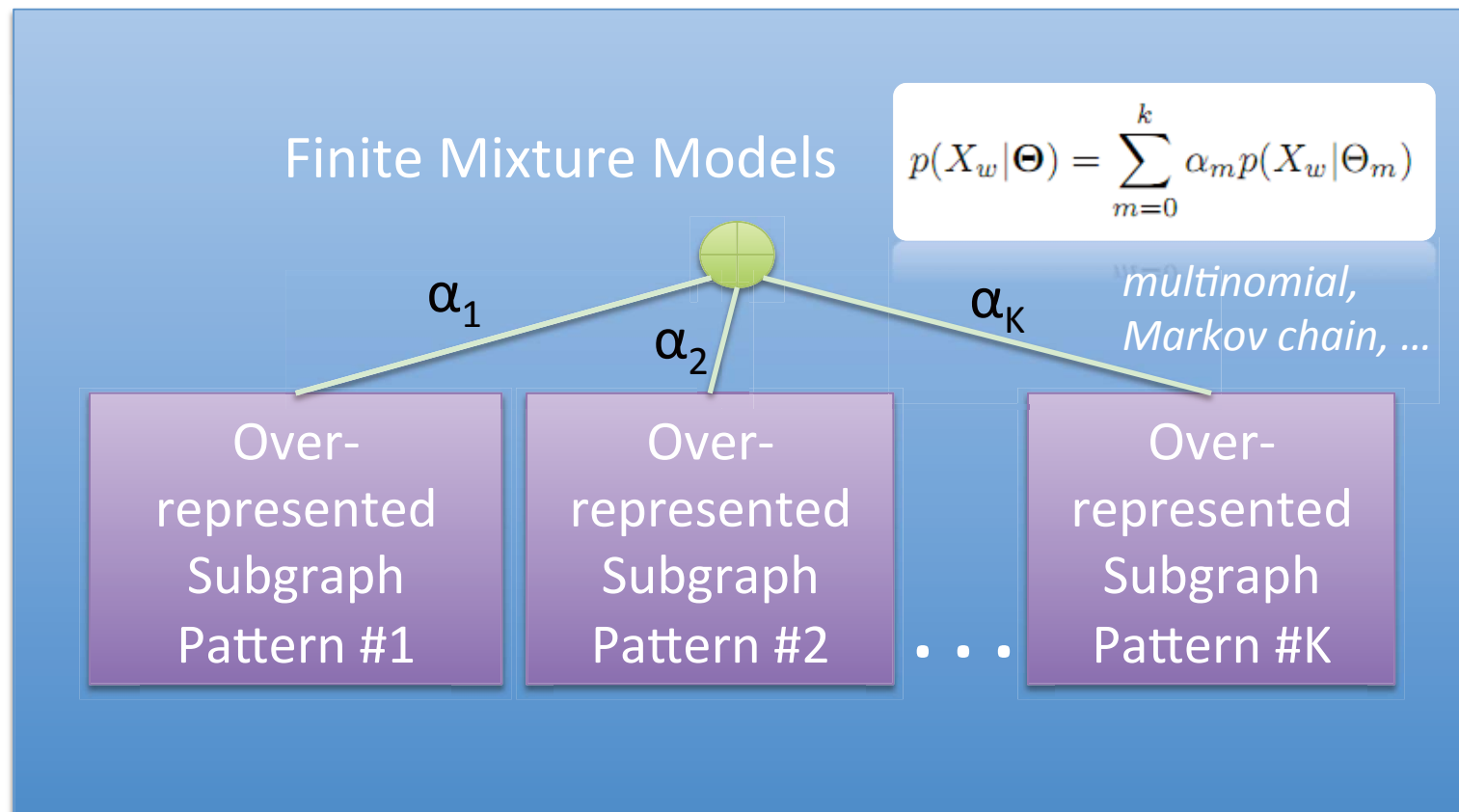
Highlights of Our Approach

- Unbiased subgraph sampling + Maximum Likelihood model parameter estimation

- Outperform deterministic network motif detection in term of *noise tolerance*

- The *optimal number of motifs* can be automatically determined by Component EM.

- Stochastic Network Motif vs. Markov Random Field
  - SNM allows *stochastic structural variations* for each network motif whereas MRF only model the distribution of cliques of different sizes.

Network Motifs in Wiki Vote Network

Wiki Vote Dataset:
- # nodes (individuals) =7115; # edges (votes) =103689
- Promoting to adminship via voting
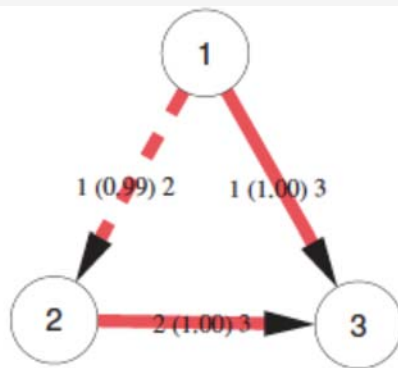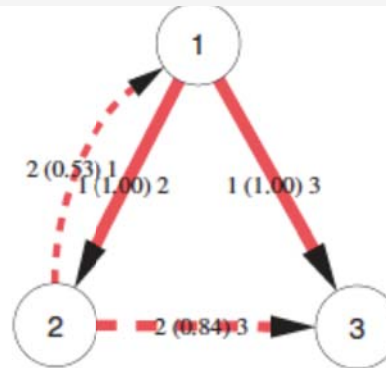
(a) Wiki1   (b) Wiki2   (c) Wiki3

If "1" votes "2" and "1" votes "3",
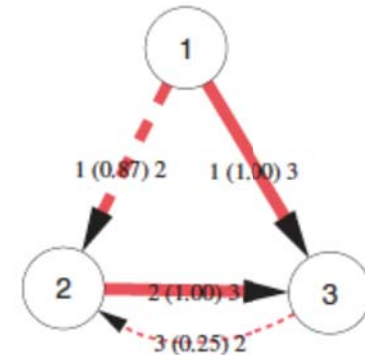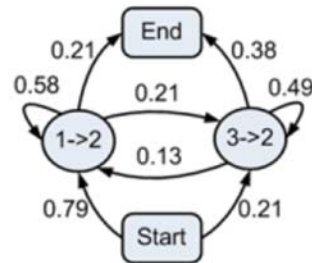if "2" should vote "3" but no vice versa.

K. Liu, W.K. Cheung, and J. Liu. Detecting multiple stochastic network motifs in network data. Advances in Knowledge Discovery and Data Mining, pages 205–217, 2012.

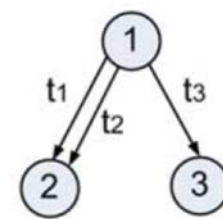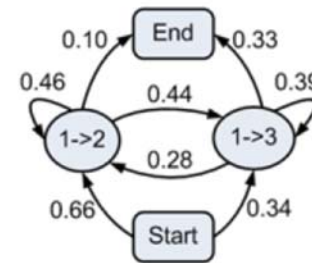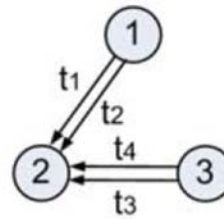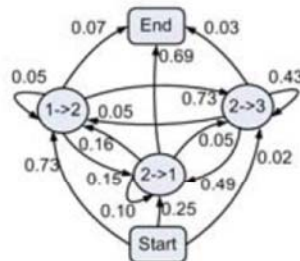# Mobile Phone Usage Motifs

Nodobo Dataset [Bell *et al.* 2011]:
- # nodes (phone numbers) =771; # edges (calls) =13035
- High school students phone calls collected in 5 months.

(a) Motif-M1

Collective action

(e) Motif-M3
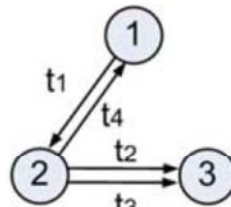
Exchange

(g) Motif-M4

Balance

K. Liu, W.K. Cheung, and J. Liu. Detecting Stochastic Temporal Network Motifs for Human Interaction Analysis. *Submitted*
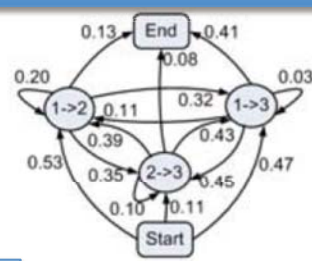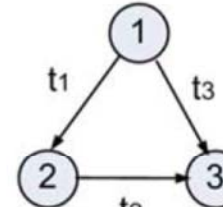
Movement Motifs in A Smart Home
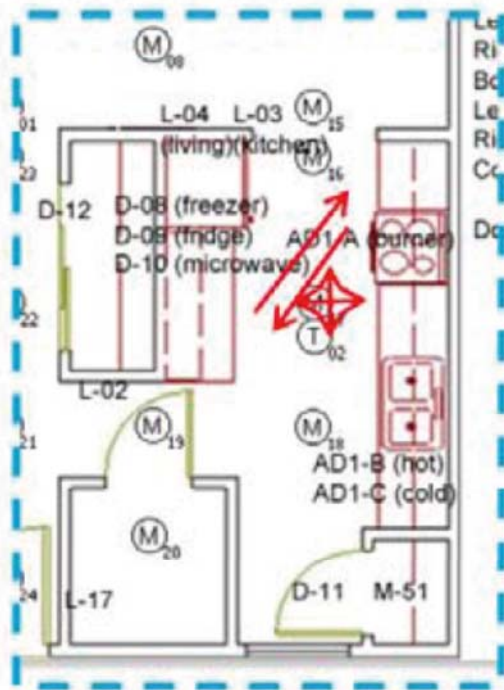
(a) Motif1 — Staying within the kitchen
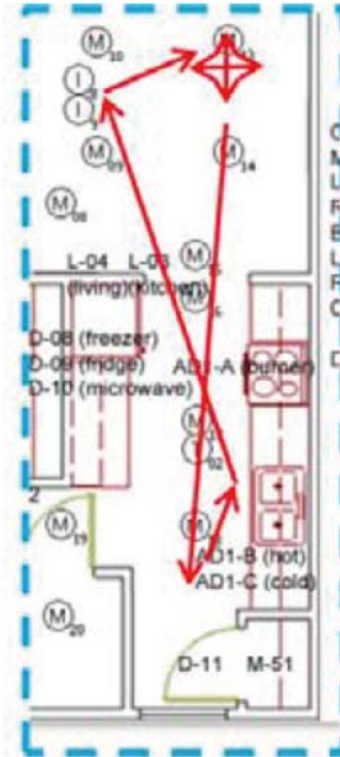
(b) Motif2 — Living rooms and kitchen

(c) Motif3 — Cooking + moving in/out of kitchen

K. Liu, W.K. Cheung, and J. Liu. Extracting behavioral motifs for characterizing human daily activities in smart environments. In Proceedings of ACM SIGKDD Workshop on Health Informatics, 2012.

Future Directions

- Multi-level network motifs and their evolution

- Heterogeneous (colored) nodes/edges
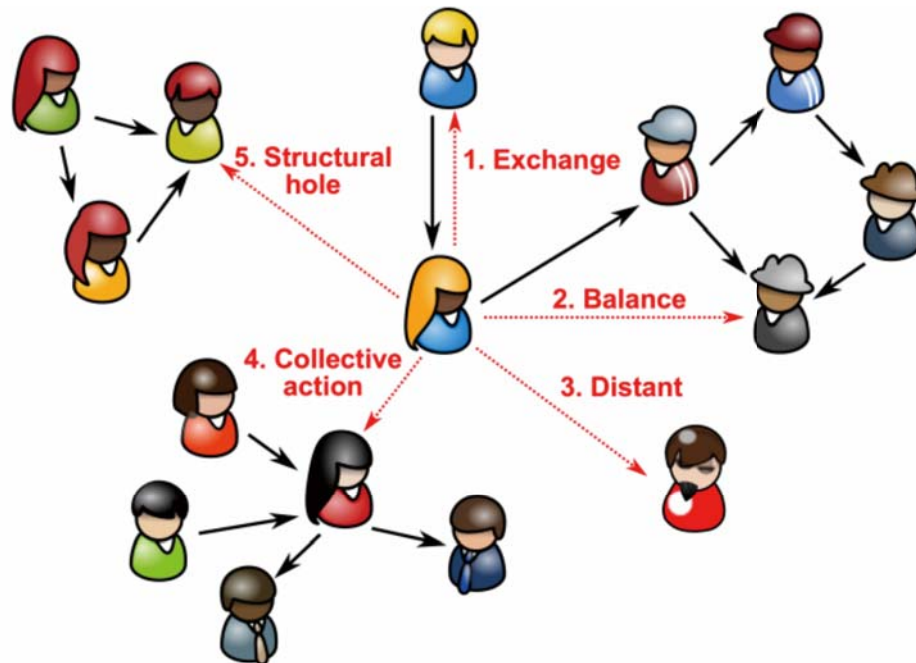
More Applications

- Elderly in nursing homes
  - changes in interaction patterns -> changes in the mental or physical status of an elderly (e.g., Dementia)

- Healthcare workers
  - contact patterns -> functional roles (hubs/bridges?) in disease spreading within hospitals / in healthcare workflow efficiency

- Youth in blogspace
  - interaction patterns -> psychological conditions (e.g. social withdrawal)

I/We do not know enough – interdisciplinary nature.

Focus 2 – Modeling Influence via Structural Diversity

Friendship neighborhood:

Contact neighborhood:

○ Connected components
○ Components of size ≥ 3

Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion, Proc Natl Acad Sci U S A. 2012 April 17; 109(16): 5962–5966.

Diffusion Network Inference

Network G^*

Cascade c_1

Cascade c_2

Cascade c_3

Node i
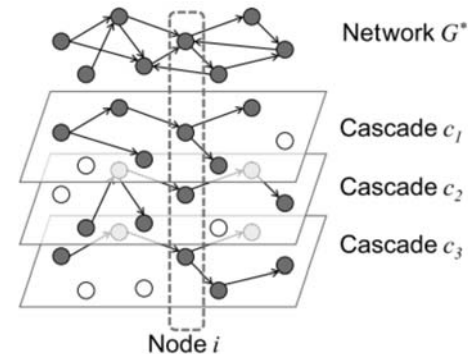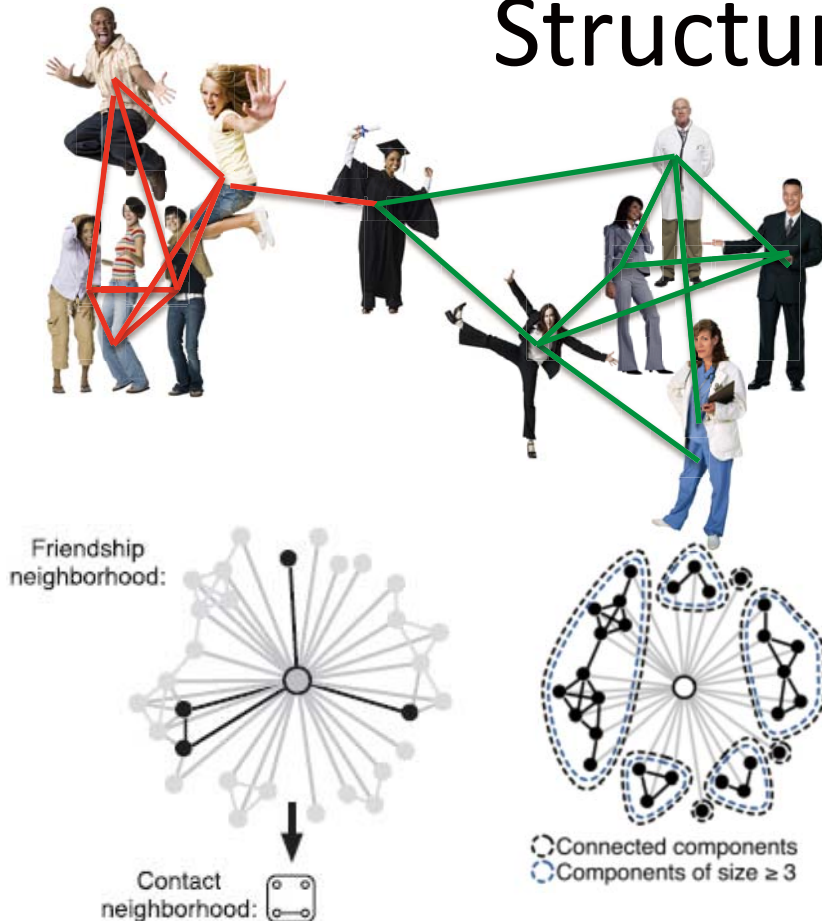
ML Estimation

Recommender Systems

bear polar	tree forest
snow| tundra	cat tiger
tundra polar	crowd bengal
fog bear ice	stream tiger cat
bear polar	cat bengal
tundra snow ice	tiger forest stream
bear tundra polar	tree cat bengal
snow ice	tiger forest

Focus 3 – Workflow mining (LI Chen)
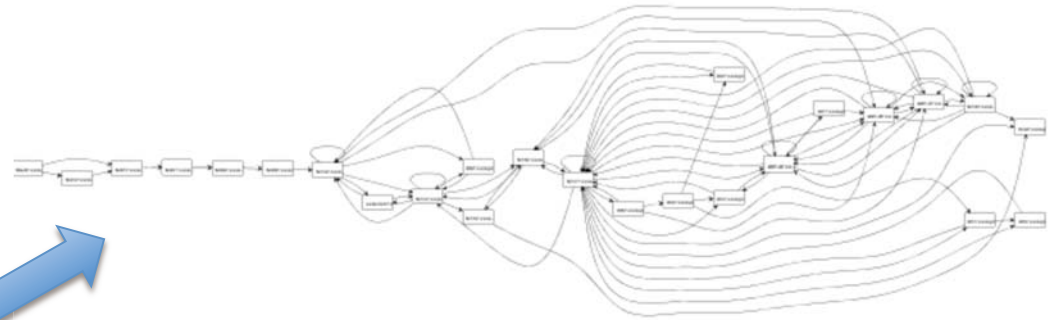
- Smart Home Dataset
- Petri Net Learning

http://www.healthcare-analytics-process-mining.org/
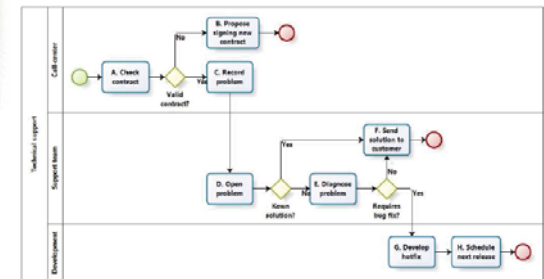
Future / Hope

- More informed decisions on complex issues, without the hindrance on usable data analysis tools.

- Take care of people who need our *better attention* by more systemic behavioral study.

- More time for *in-depth* chat, food and drink, … and innovations.