Biologically Inspired Community Detection

Zohar Nussinov

WASHINGTON UNIVERSITY, ST. LOUIS
The problem

How may we detect pertinent structures in general complex biological and other systems *on all scales*? How can we engineer systems inspired by cooperative effects in biology and sociology?

Possible tool:

“Multi-resolution community detection”

*No Guess-work or assumptions!*
A small section of a biological system. Each system comprises a convoluted network of smaller pathways, each with its own level of complexity. Image courtesy of Merck Research Laboratories.

SYSTEMS BIOLOGY:

Studying the World’s Most Complex Dynamic Systems

BY RICARDO PAXSON AND KRISTEN ZANNELLA
Key idea:

Use inspiration from biology itself-employ (information theory) correlations between different solvers that minimize a global energy sum to find an optimal partition into different groups (or “communities”) on all scales.
What is “community detection”?

Community Detection

- Node – basic element of abstracted graph
- Edge – a defined relationship between two nodes
  - unweighted or weighted
  - directed or undirected
- Community – nodes are more densely connected within their own community than to other communities
Community Detection with a Potts Model

- Identify Nodes as Potts model spins
- Spin orientation $\sigma \leftrightarrow$ community assignment (with ferromagnetic and anti-ferromagnetic interactions)
- Energy Contributions
  - Inside Connected ($-$)
  - Inside Unconnected ($+$)
  - Outside Connected ($+$)
  - Outside Unconnected ($-$)
Community Detection with a Potts Model

- Simplifies to

\[ H(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} (a_{ij}A_{ij} - b_{ij}J_{ij}) \delta(\sigma_i, \sigma_j) \]

- ‘good’ contribution: inside connected
  - Sum only over nodes inside communities
  - Significantly improves algorithm properties and performance

- ‘bad’ contribution: inside unconnected

- Uses a global energy sum...
  - but it is a local community measure

- In practice, we search for low energy states
Community Detection with a Potts Model

- Insert a model weight $\gamma$

\[
H(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} (a_{ij} A_{ij} - \gamma b_{ij} J_{ij}) \delta(\sigma_i, \sigma_j)
\]

- $\gamma$ allows the user to set the system scale – equivalent to the system “resolution”

- $\gamma \leftrightarrow \text{minimum community density } p_i$ for every community

\[
p_i = \frac{\gamma}{\gamma + 1}
\]
Multiresolution Community Detection

- Systems can differ in optimal divisions based on the scale at which the system is examined
  - Social relationships
    - Family
    - Close friends
    - Friends
    - Acquaintances
  - Others
    - Internet (not router)
    - Language
    - Air travel network
Multiresolution Systems

- Types of multiresolution structure:
  - Hierarchical
  - Overlapping structures (nodes shift memberships at different scales)
  - Other multiresolution structures?
Multiresolution Issues: Selecting the “best” scale...

- As a local measure, the APM does not ‘scale’ with the system (like those that use null models)
- How do we choose the ‘best’ scale(s)?
Multiresolution Method

- ‘Replica’ – Independent solution of the same system at the same model weight $\gamma$

- Solve multiple replicas over a range of “resolutions” (by varying $\gamma$)

\[
H(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} (a_{ij} A_{ij} - \gamma b_{ij} J_{ij}) \delta(\sigma_i, \sigma_j)
\]

- Pick the ‘best’ resolutions based how strongly the replicas are correlated
Multi-Resolution Method

- Strong correlations (using NMI or VI) between replicas indicate the ‘best’ resolutions.

- Corresponding NMI or VI value is a quantitative estimate of the strength of each division.

- This quantitative estimate is missing in other multiresolution community detection methods.
Multi-Resolution Test System: Hierarchical

- Test system is a heterogeneous 3-level hierarchy
  - Level 1: 256 nodes with a random density of 0.1 between sub-communities
  - Level 2: 5 groups of sizes from 33 to 79 with an edge density of 0.3 each between sub-communities
  - Level 3: 16 groups of sizes from 5 to 22 with an interior edge density of 0.9 each
Multiresolution Test System

- \((ia,b) \rightarrow \)
  Level 2: 5 communities
- \((iia,b) \rightarrow \)
  Level 3: 16 communities
- VI peaks \(\rightarrow\) maximum "complexity" (zero energy difference between states)
Known applications

Protein networks
Food networks
Communication networks
Social networks

....
Applications (non-network
Artificial vision (no training sets)

- **Generalized Potts Model**

\[ H(\sigma) = \frac{1}{2} \sum_{a=1}^{q} \sum_{i,j \in C_a} [(V_{ij} - \overline{V})[\Theta(\overline{V} - V_{ij}) + \gamma \Theta(V_{ij} - \overline{V})]] \]

- \( \overline{V} \) - background
- \( V_{ij} \) - edge weight
Image segmentation ↔ community detection

• i) Pixels in an image ↔ nodes in a graph

• ii) The color/intensity similarity ↔ Edge weights
Image segmentation results
Image segmentation results

stripes

spots

stripes
(a) The variation of information $V$ as a function of length $\ell$ at $\gamma = 0.1$

(b) The normalized mutual information $I_N$ as a function of length $\ell$ at $\gamma = 0.1$

(d) The variation of information $V$ as a function of length $\ell$ at $\gamma = 0.05$

(e) The normalized mutual information $I_N$ as a function of length $\ell$ at $\gamma = 0.05$

$l=0.63$

$l=1$

$l=1.29$
Determination
Of optimal parameters
(a) The normalized mutual information $I_N$ as a function of the resolution $\log(\gamma)$ and temperature $T$. 
Chemical structures: amorphous
Low T
ZrPt glass
Al$_{88}$Y$_7$Fe$_5$ at a temperature of $T = 300$K. The panels at left (a,b) show the information theoretic overlaps between the different replicas when averaged over all replica pairs (see text). These are the variation of information (VI), mutual information (I), normalized mutual information (NMI), entropy (H) and number of clusters (q) in individual paritions. On the right (c), we highlight the corresponding spatial structures. The clusters found are assigned different colors.
The corresponding structure of $\text{Al}_{88}\text{Y}_{7}\text{Fe}_{5}$ at a temperature of $T=1500\text{K}$.
Collaborators:

P. Ronhovde
(Phys. Rev. E 80, 016109 (2009);
PRE 81, 046114 (2010))

D. Hu
(arXiv:1008.2699, 1106.5793)

S. Chakrabarty,
K. K. Sahu
M. Sahu
K. F. Kelton
N. A. Mauro
(arXiv: 1101.0008, 1102.1519)