A Spin-Glass Model for Semi-Supervised Community Detection

Eric Eaton
Bryn Mawr College

Rachael Mansbach
University of Illinois at Urbana-Champaign

Abstract

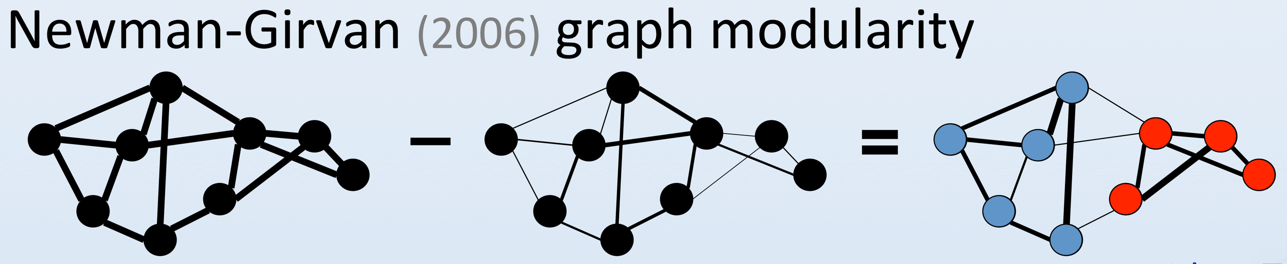
Current modularity-based community detection methods show decreased performance as relational networks become increasingly noisy. These methods also yield a large number of diverse community structures as solutions, which is problematic for applications that impose constraints on the acceptable solutions or in cases where the user is focused on specific communities of interest. To address both of these problems, we develop a semi-supervised spin-glass model that enables current community detection methods to incorporate background knowledge in the forms of individual labels and pairwise constraints. Unlike current methods, our approach shows robust performance in the presence of noise in the relational network, and the ability to guide the discovery process toward specific community structures. We evaluate our algorithm on several benchmark networks and a new political sentiment network representing cooperative events between nations that was mined from news articles over six years.

Introduction

- Newman-Girvan graph modularity (Newman 2006) is the foundation for many automatic community detection methods
- Current modularity-based methods exhibit two key problems:
  1. the inability to handle noise in the network
  2. the tendency to admit a large number of high-scoring solutions without a clear optimum (Good et al. 2010)
- We incorporate user guidance into the community detection process to:
  - augment its performance in noisy networks
  - focus discovery on communities of interest to the user

Background on Newman-Girvan Graph Modularity

- Relational network given by $G = (V, \mathbf{A})$
  V: set of n vertices $\quad \mathbf{A}$: n x n adjacency matrix, m total edges
- Newman-Girvan (2006) graph modularity



- Measures the global community structure of a partitioning C:

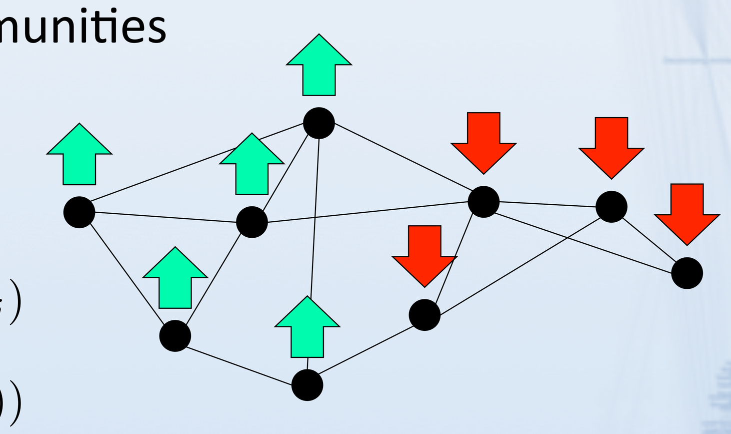
$$Q(C) = \frac{1}{2m}\sum_{i,j}(A_{ij} - P_{ij})\underbrace{\delta(C_i, C_j)}_{\text{Kronecker delta}} \qquad P_{ij} = \frac{d_i d_j}{2m}$$

Kronecker delta — community of vertex i

From Modularity to Spin-Glass Models

- Graph modularity is a special case of the Potts spin-glass model from statistical mechanics (Reichardt & Bornholdt, 2006)
  - Ground state corresponds to communities
  - Found by minimizing Hamiltonian:

$$\begin{aligned}\mathcal{H}(C) = &-\sum_{i\neq j} a_{ij}A_{ij}\delta(C_i,C_j)\\ &+\sum_{i\neq j} b_{ij}(1-A_{ij})\delta(C_i,C_j)\\ &+\sum_{i\neq j} c_{ij}A_{ij}(1-\delta(C_i,C_j))\\ &-\sum_{i\neq j} d_{ij}(1-A_{ij})(1-\delta(C_i,C_j))\end{aligned}$$



  - Choosing the coefficients $\{a, b, c, d\}$ appropriately, we can recover Newman-Girvan modularity

Incorporating User Guidance

- Idea: Penalize for communities that violate the guidance
- Penalty measured by a function $U : C \mapsto \mathbb{R}$ of the form

$$U(C) = \sum_{i\neq j}\Big(\underbrace{u_{ij}}_{\text{same community penalty}}\big(1-\delta(C_i,C_j)\big) + \underbrace{\overline{u}_{ij}}_{\text{different communities penalty}}\delta(C_i,C_j)\Big)$$

- Guidance can be incorporated directly into the Hamiltonian as

$$\mathcal{H}'(C) = \underbrace{\mathcal{H}(C)}_{\text{original Hamiltonian}} + \mu\sum_{i\neq j}\Big(u_{ij} - (u_{ij} - \overline{u}_{ij})\delta(C_i,C_j)\Big)$$

- Re-deriving graph modularity yields

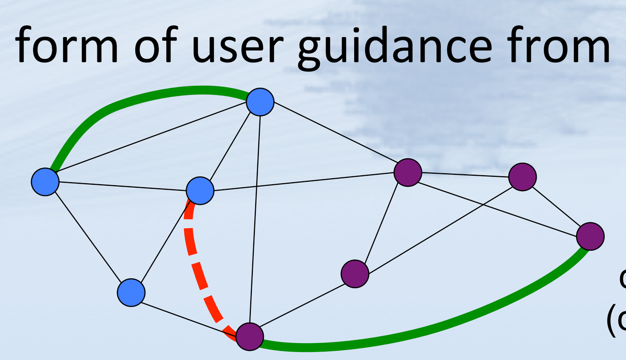
$$Q'(C) = \frac{1}{2m}\sum_{i\neq j}\left(A_{ij} - \underbrace{\left(\frac{d_i d_j}{2m} - \mu(u_{ij} - \overline{u}_{ij})\right)}_{\text{new null model}}\right)\delta(C_i,C_j)$$

Forms of Guidance

- Individual labels: same labels imply same community

$$u_{ij} = \begin{cases} 1 & \text{when } \mathrm{label}_i = \mathrm{label}_j \neq UNKNOWN \\ 0 & \text{otherwise}\end{cases} \qquad \overline{u}_{ij} = 0$$

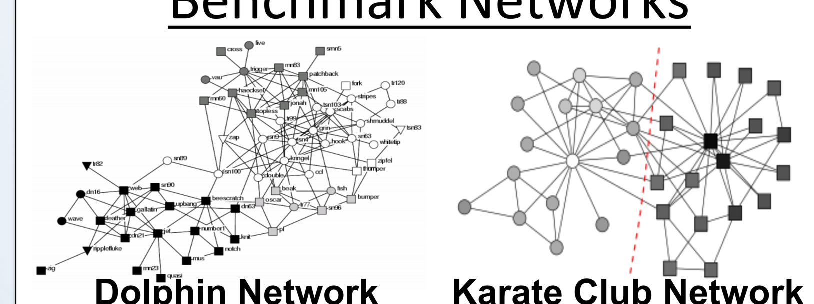
- Pairwise must-link or cannot-link constraints (Wagstaff et al. 2001)
  - Intuitive form of user guidance from constrained clustering



  - Penalties given by $u_{ij} = \underbrace{\alpha_1}_{\text{must-link weight}} w_{ij} \qquad \overline{u}_{ij} = \underbrace{\alpha_2}_{\text{cannot-link weight}}\overline{w}_{ij}$

Results

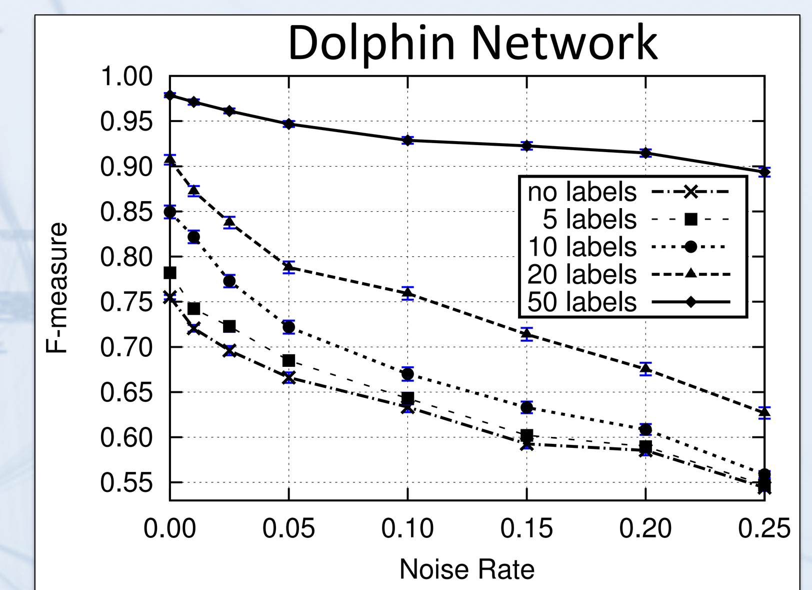
Benchmark Networks

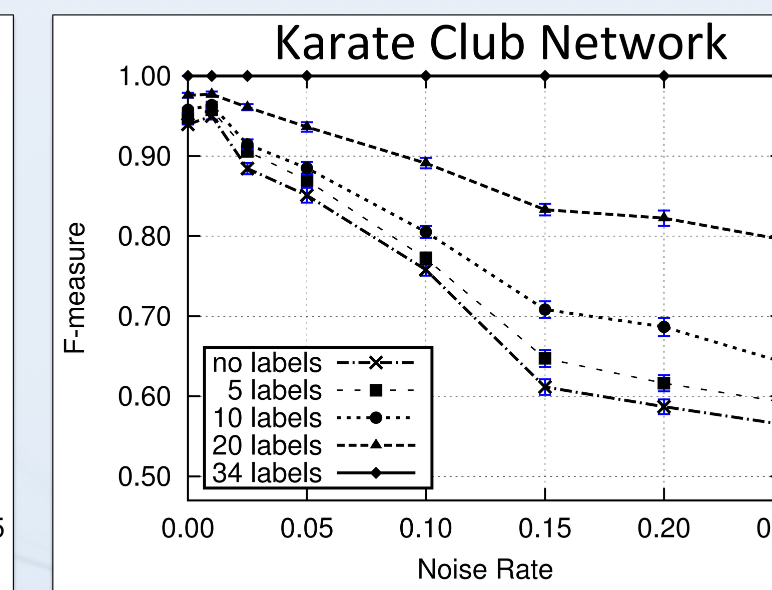


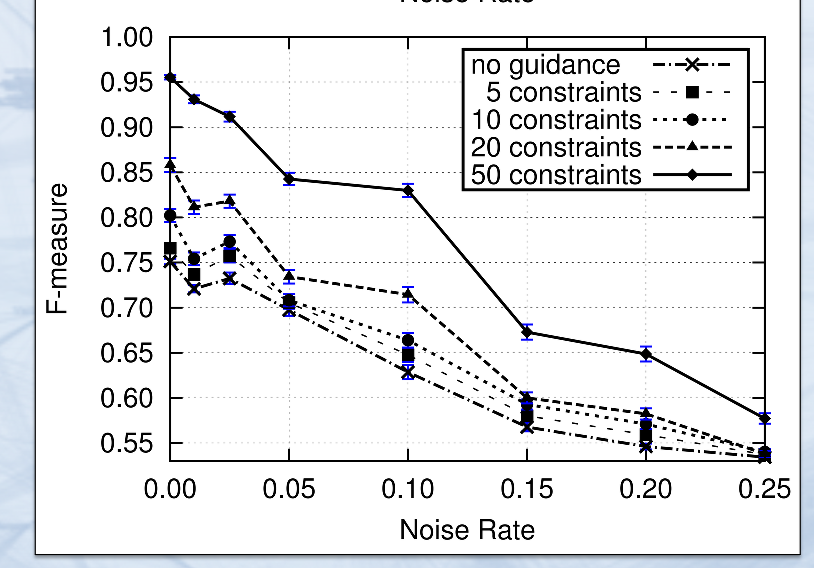
Application Network

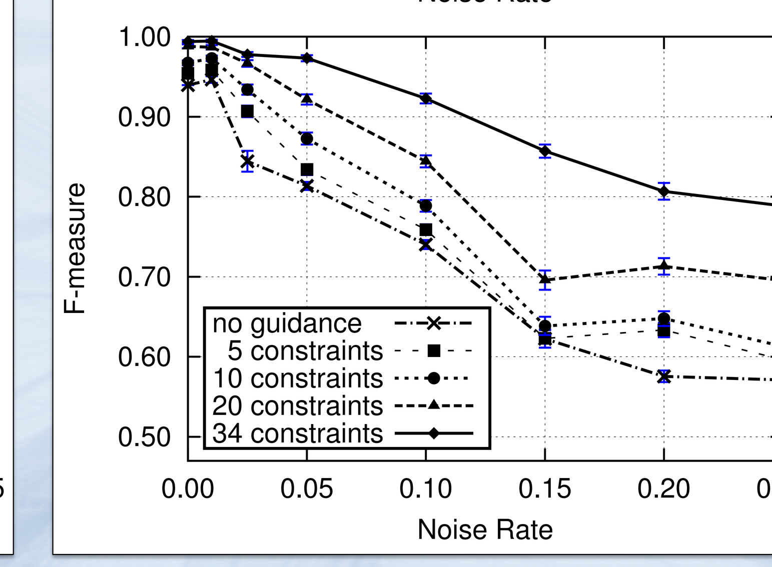
- Political network measuring cooperative/hostile acts between nations
  - 336,555 events between 196 nations from Jan. 2005 – Dec. 2010

Improved performance on noisy networks

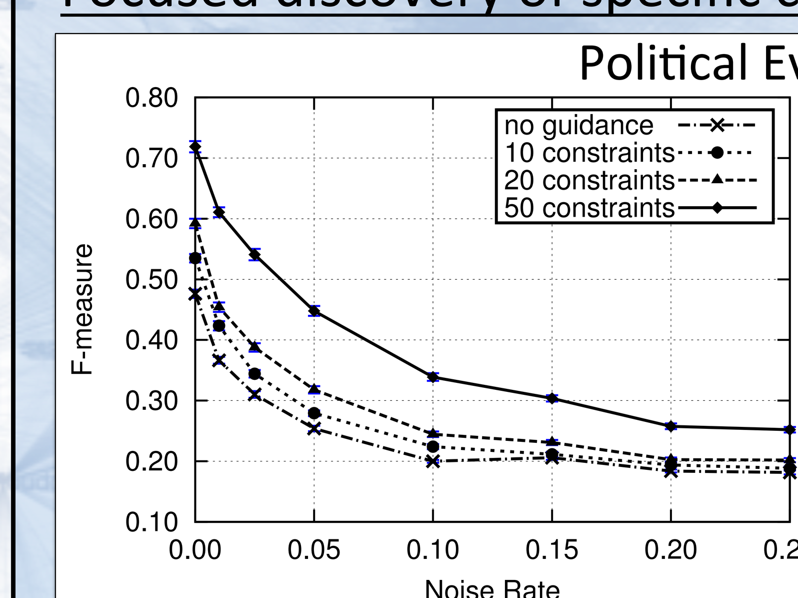


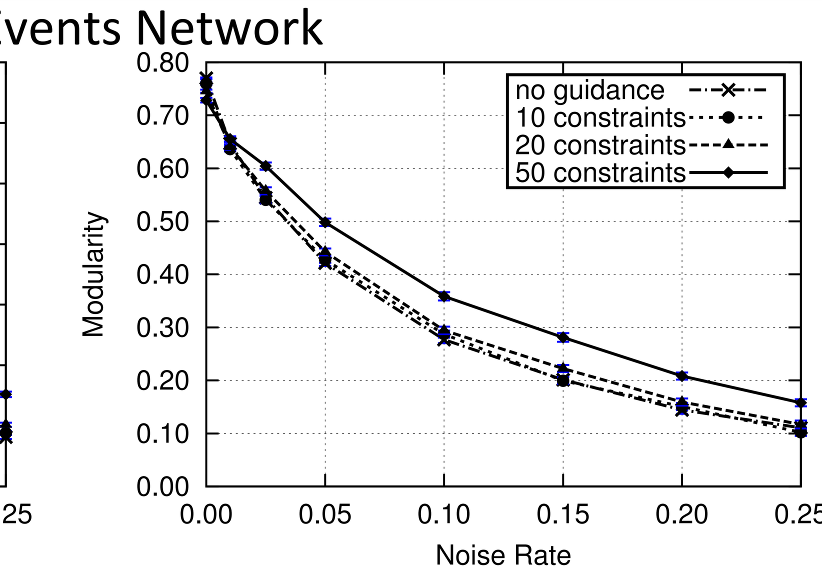




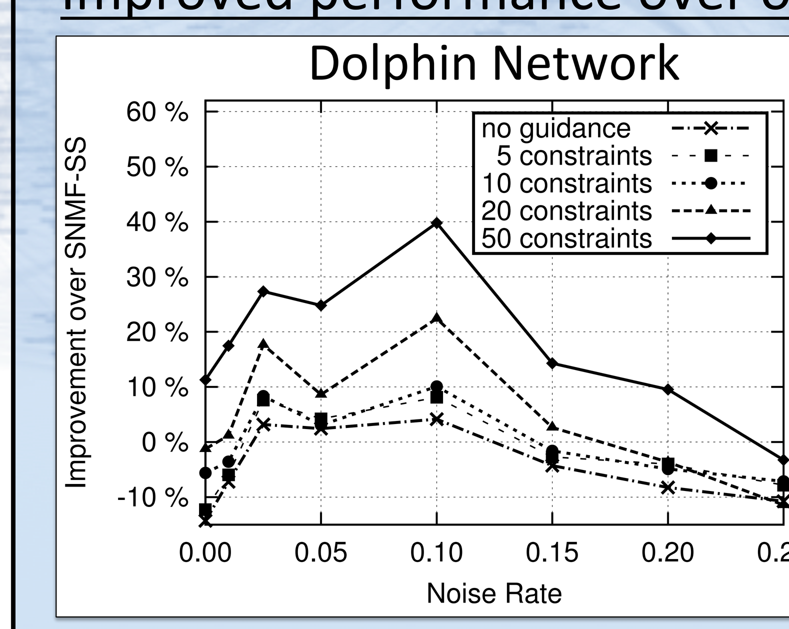


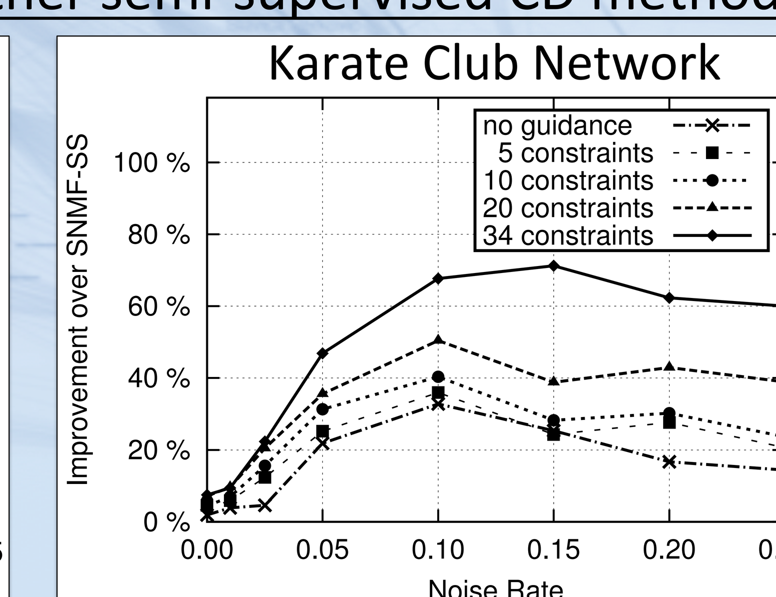
Focused discovery of specific communities of interest





Improved performance over other semi-supervised CD methods





This research was supported by ONR grant #N00014-10-C-0192