GLAD: Group Anomaly Detection in Social Media Analysis

Poster #: 1150

Rose Yu, Xinran He and Yan Liu

University of Southern California

Group Anomaly Detection

Anomalous phenomenon in social media data may not only appear as individual points, but also as **groups**.

Group Review Spamming

Organized Viral Campaign

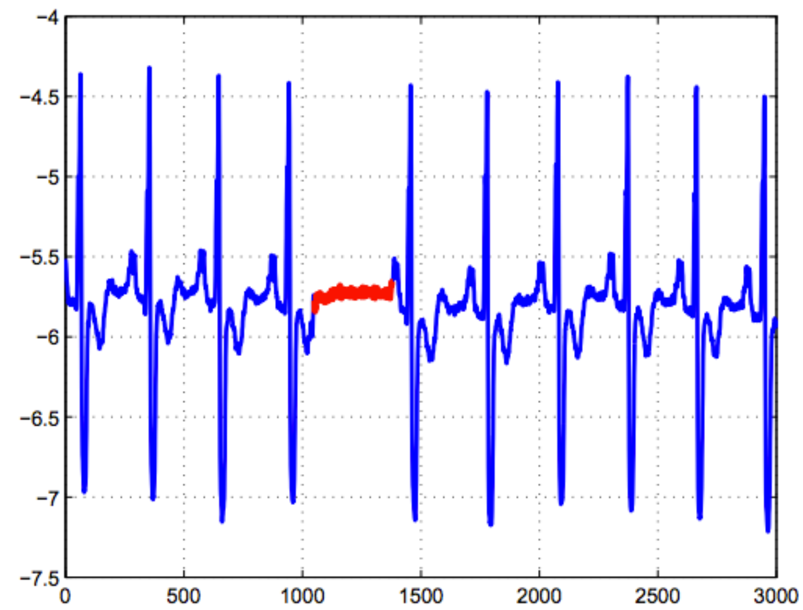
Massive Cyber Attack

We develop a hierarchical Bayes model to automatically discover social media groups and detect group anomalies simultaneously.

Challenges

Detecting group anomalies require us to exploit the structure of the social media groups as well as the attributes of the individual points.

- Point-wise features data and pair-wise relational data coexist and are mutually dependent.
- Collective anomalous activities might appear normal at the individual level. [*Chandola2007*]
- People constantly switch groups and we can hardly know groups beforehand.



[An Example of Collective Anomaly]

Group Anomaly Definition

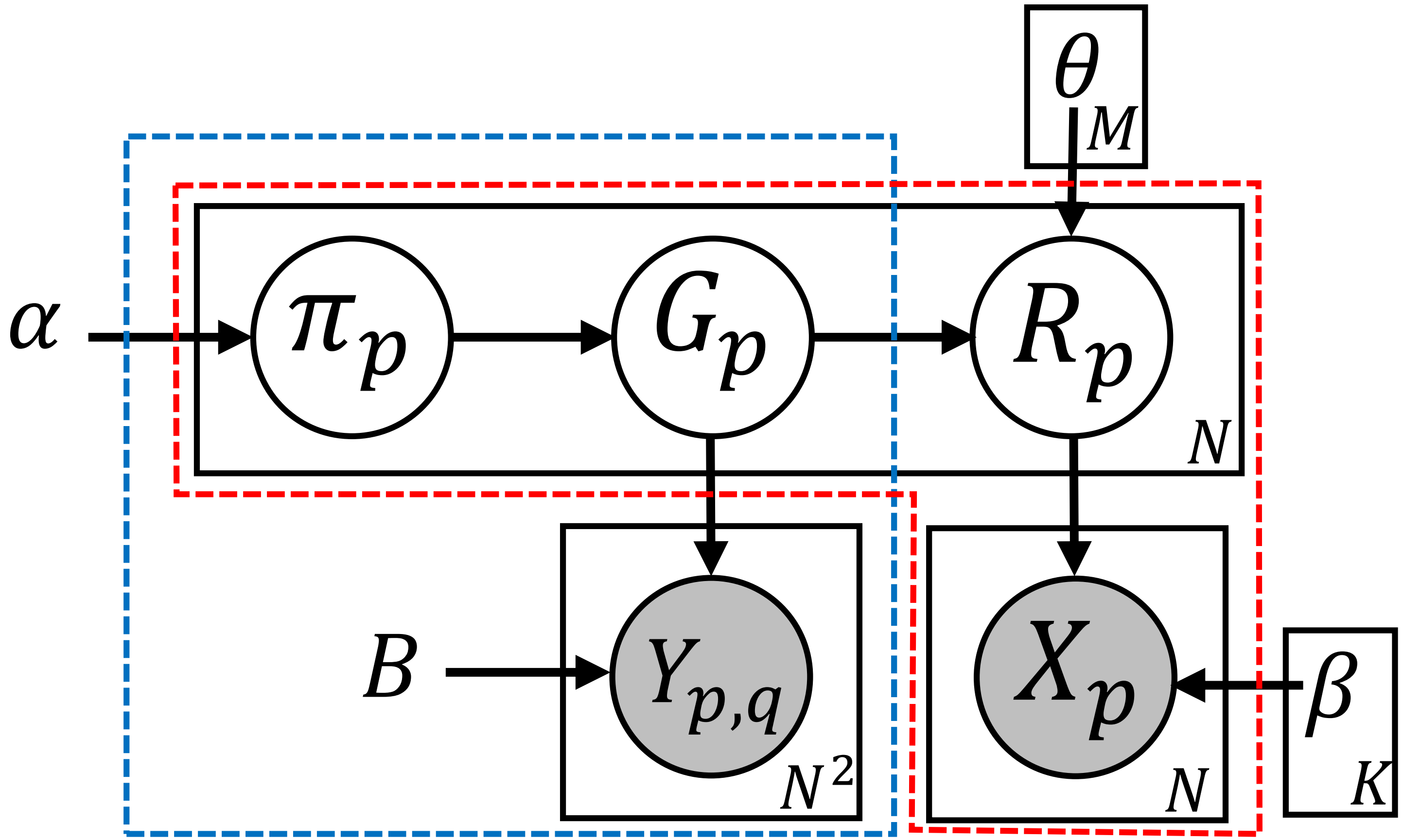
- Input:
  - Pairwise communication network;
  - Pointwise node features;
- Output:
  - List of groups ranked by anomaly score
- Model a group as a mixture of roles
  - Same roles, different role mixture rate

Definition: Group anomaly has a *role mixture rate* pattern that does not conform to the majority of other groups.

- Existing Approaches:
  - Mixture Genre Model (MGM) [xiong et al 2011a]
  - Flexible Genre Model (FGM) [xiong et al 2011b]
  - Support Measure Machine (SMM) [muandet et al 2013]

Two- Stage Approaches!

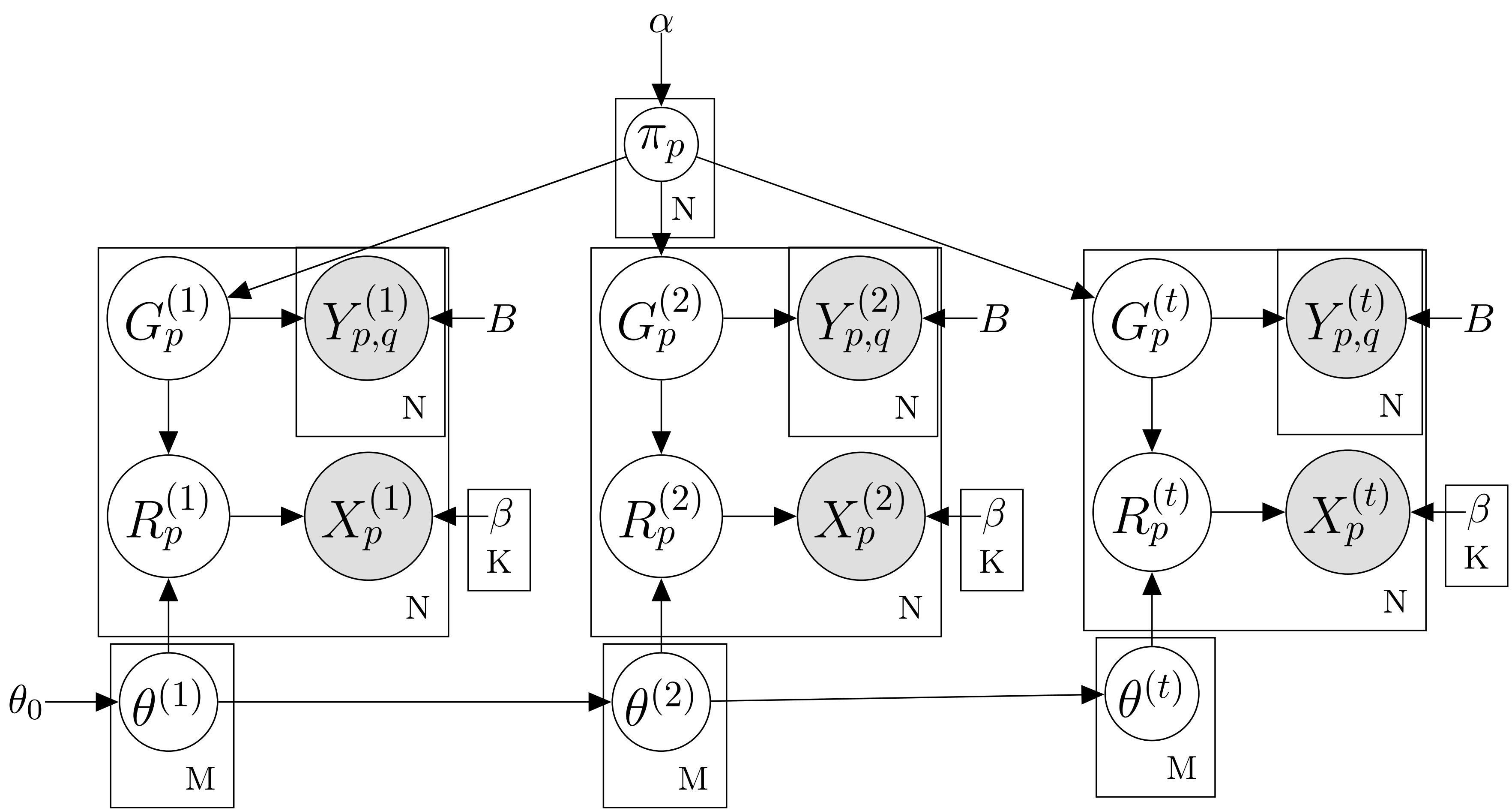
Group Latent Anomaly Detection (GLAD)



$\pi_p \sim Dirichlet(\alpha)$ $\quad G_p \sim Categorical(\pi_p)$ $\quad R_p \sim Categorical(\theta_{G_p})$

$Y_p \sim Bernouli(B_{G_p,G_q})$ $\quad X_p \sim Multinomial(\beta_{R_p})$

Dynamic extension of GLAD (d-GLAD)



$$\theta_m^t \sim Gaussian(\theta_m^{t-1}, \sigma)$$

Procedure

Use BIC to decide # of groups and # of roles → Learn GLAD/ d-GLAD to infer role mixture rates → Rank groups with respect to the anomaly score → Perform significant test and raise alarms

- GLAD : the expected likelihood of role distribution

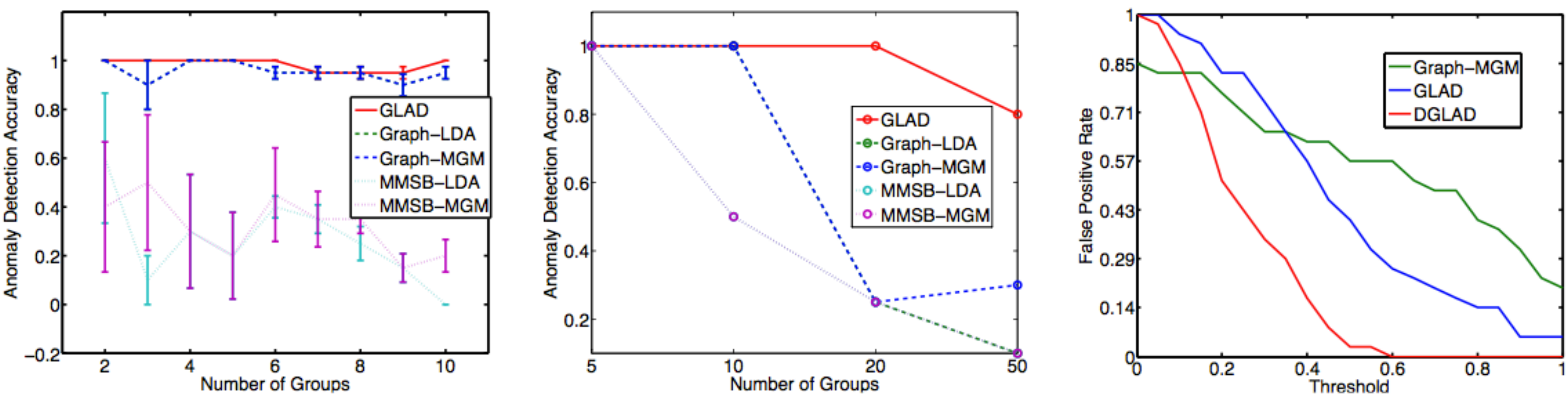
$$AnomalyScore_{GLAD} \sim -\sum_{p \in G} E_q[p(R_p \mid \theta)]$$

- d-GLAD : the change of role mixture rate over time

$$AnomalyScore_{d-GLAD} \sim \| \theta_m^{t-1} - \theta_m^{t} \|_2$$

Synthetic Experiments

- 500 nodes network, 2 roles, different number of groups
- Normal group mixture rate [0.9, 0.1], anomalous group mixture rate [0.1,0.9] , 20% injected group anomaly



(a) Robustness (b) Accuracy (c) d-GLAD

Detect ``anomalous `` research communities

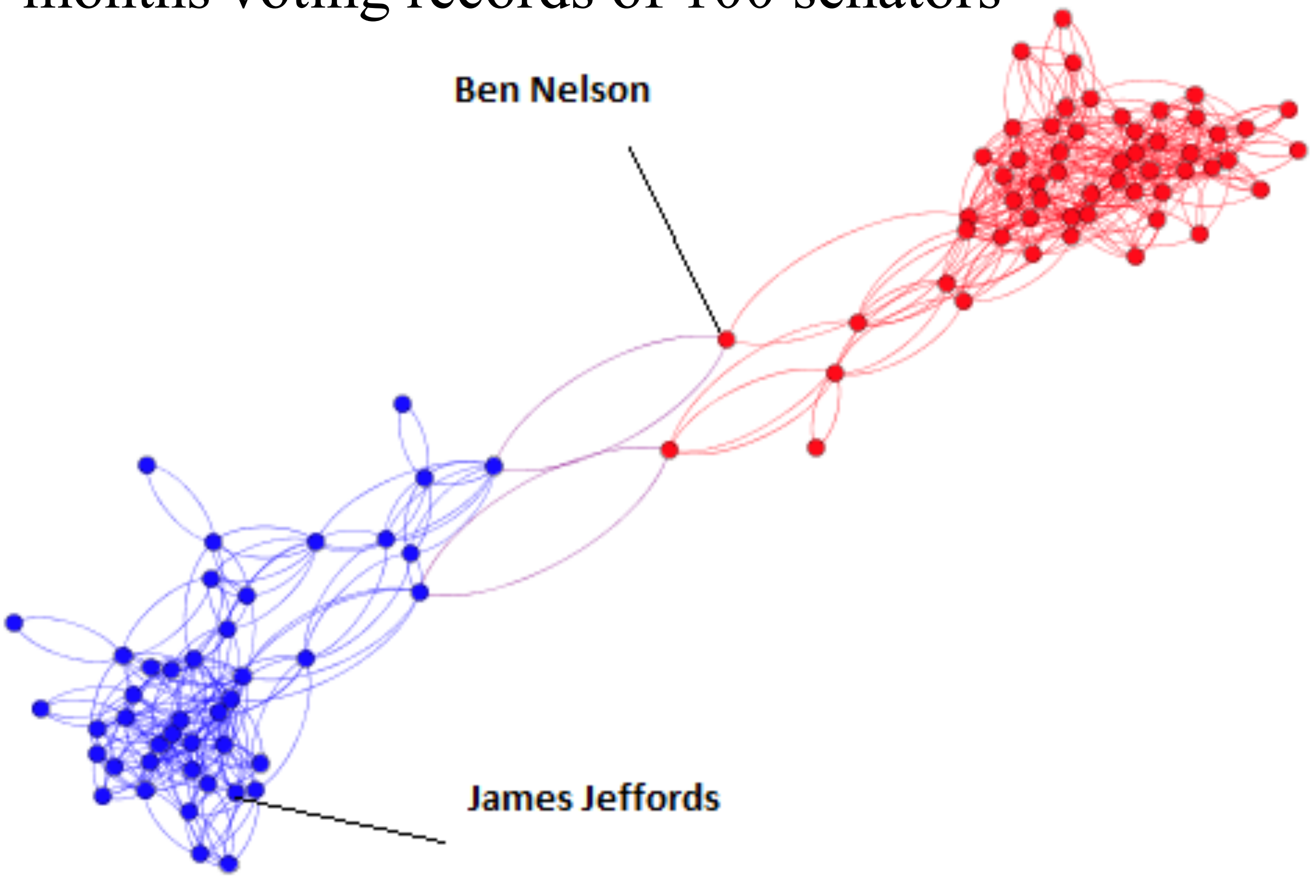
- Pointwise: Bag of words of paper abstracts
- Pairwise: Common authorship
- 28,702 authors, 104,962 links, 11,771 vocabulary size, 4 topics. Injected 20% group anomalies from other venues.

Table : Accuracy of detecting other venues from KDD paper groups.

Methods	GLAD	Graph-LDA	Graph-MGM	MMSB-LDA	MMSB-MGM
KDD/CVPR	**0.4167**	0.3333	0.3333	0.2500	0.2500
KDD/ICML	**0.2500**	0.0833	0.0833	0.1667	0.1667
KDD/SIGMOD	**0.2875**	0.0750	0.0500	0.1625	0.1625
KDD/EDBT	**0.2625**	0.0500	0.0875	0.2000	0.2000

Detect Party Affiliation Changes

- Pointwise: Senator attributes
- Pairwise: Common votes
- 24 months voting records of 100 senators



Conclusion

- Formulate the problem of group anomaly detection social media analysis for both static and dynamic settings
- Develop a unified hierarchical Bayes model GLAD to infer the groups and detect group anomalies simultaneously
- Experiments on both synthetic, academic publications and senator voting datasets show benefits over two-stage approaches .

Future Work

- Non-parametric Bayes model to automatically decide # of groups and # of roles.
- Better detection evaluation procedures and metrics other than anomaly injection.

References

- [1] V. Chandola, A. Banerjee, and V. Kumar. Outlier detection: A survey. 2007.
- [2] Liang Xiong, Barnab´as P´oczos, Jeff G Schneider, Andrew Connolly, and Jake Vanderplas. Hierarchical probabilistic models for group anomaly detection. In *International Conference on Artificial Intelligence and Statistics*, pages 789–797, 2011.
- [3] Liang Xiong, Barnab´as P´oczos, and Jeff G Schneider. Group anomaly detection using flexible genre models. In *Advances in Neural Information Processing Systems*, pages 1071–1079, 2011.
- [4] Krikamol Muandet and Bernhard Sch¨olkopf. One-class support measure machines for group anomaly detection. *arXiv preprint arXiv:1303.0309, 2013*.
- [5] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- [6] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(1981-2014):3, 2008.

Thank You !