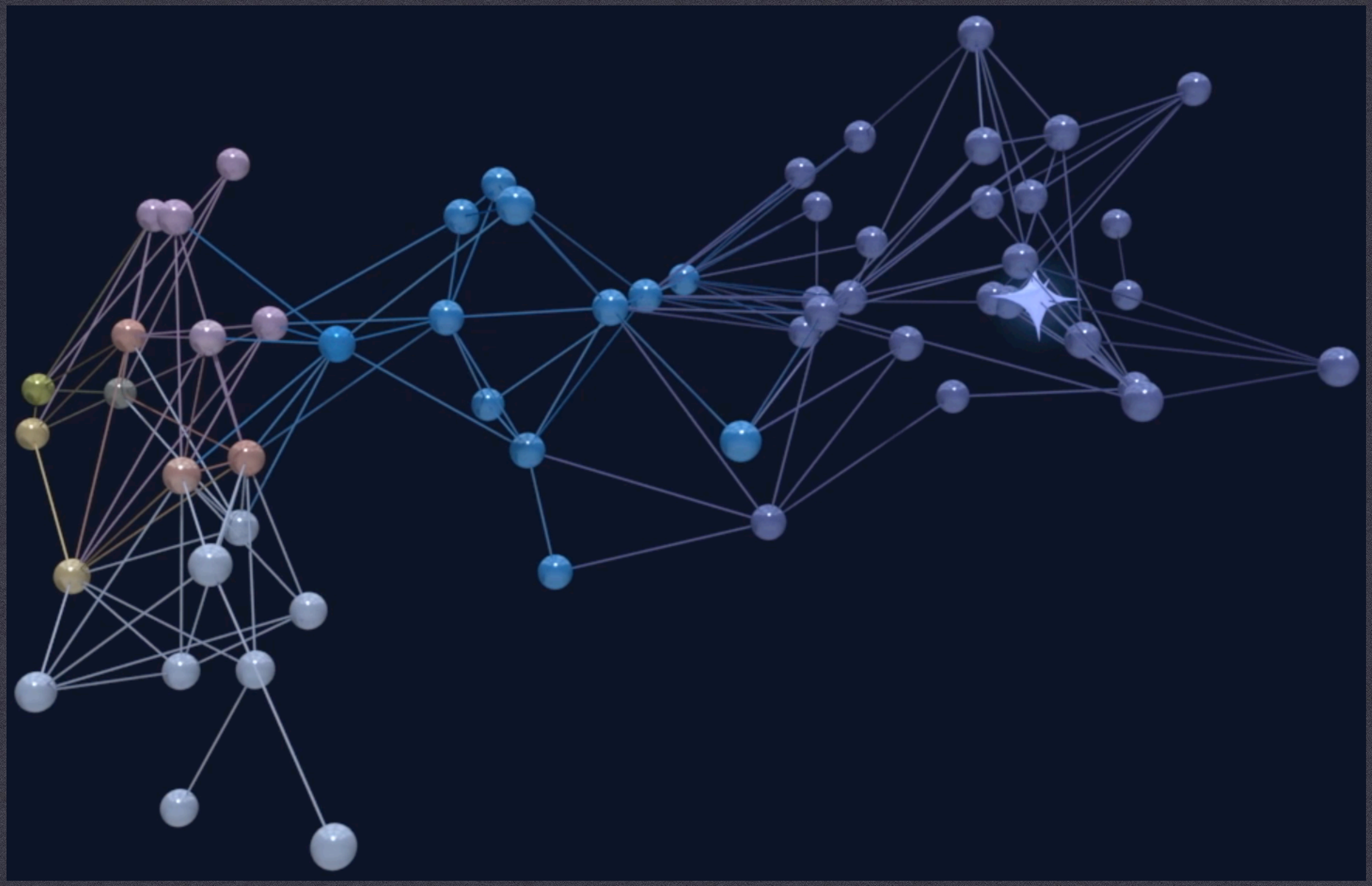


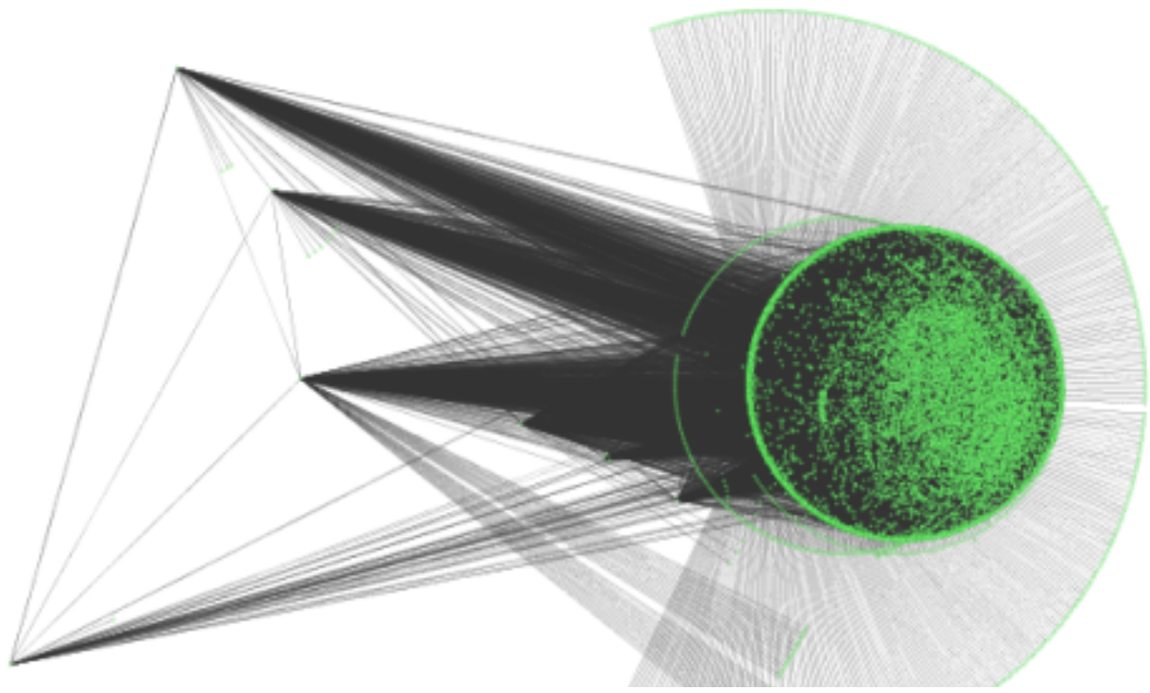
ANALYZING MASSIVE GRAPHS - PART II

DR. MICHAEL FIRE

Synergy between Networks and Machine Learning

We can use networks and machine learning to tackle a wide variety of challenges:

- Predict future links
- Predict vertex's attributes
- Create recommendations
- Identify fake profiles/bots
- Identify anomalies



Link Prediction

We can use the network's topological structure to predict missing/future links. This can be useful for a variety of tasks:

- Recommend other users/items that may be interesting to a user
- Predict future interactions (for example if user A will send message to user B)
- Identify missing parts of data
- Getting a job at Facebook

Link Features

$$\Gamma(v) := \{u | (u, v) \in \boldsymbol{E} \text{ or } (v, u) \in \boldsymbol{E}\}$$

$$\textit{Friends-measure}(u, v) = \sum_{x \in \Gamma(u)} \sum_{y \in \Gamma(v)} \delta(x, y),$$

$$\textit{total-friends}(u, v) = |\Gamma(u) \cup \Gamma(v)|.$$

$$\textit{jaccard's-coefficient}(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}.$$

$$\textit{common-friends}(u, v) = |\Gamma(v) \cap \Gamma(u)|.$$

$$\textit{total-friends}(u, v) = |\Gamma(u) \cup \Gamma(v)|.$$

$$\textit{preferential-attachment-score}(u, v) = |\Gamma(u)| \cdot |\Gamma(v)|.$$

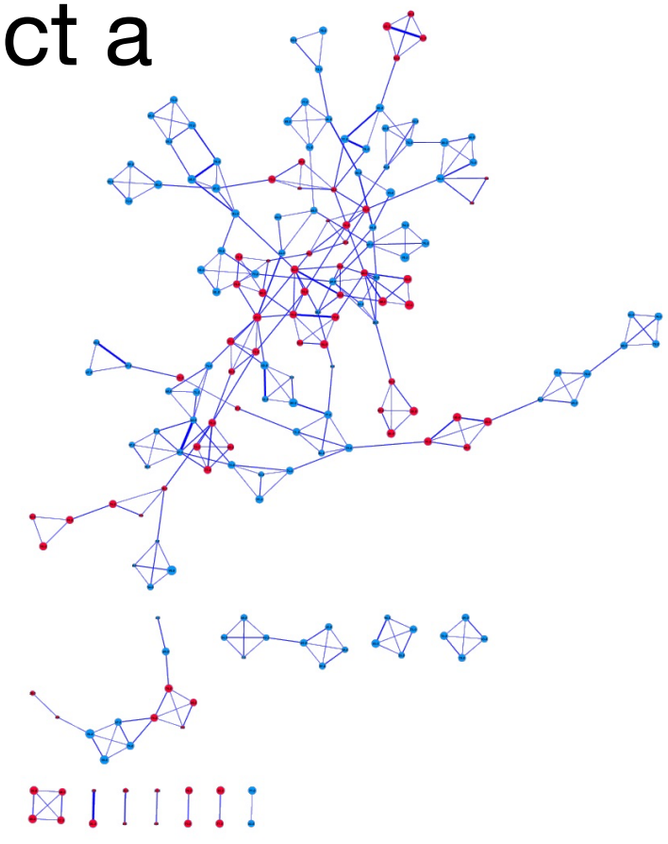
$$d(v) = |\Gamma(v)|$$

Adamic-Adar index of u and v is defined as

$$\sum_{w \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log |\Gamma(w)|}$$

Attribute Prediction

We can use networks and supervised learning to predict a vertex's attribute. For example, in the student's graph below, we are able to predict a student's test score based on his or her friends' test scores.



Network Embedding

Network embedding encompasses a variety of methods for learning feature representations of vertices, links, and subgraphs in a network. Typically, embedding methods are based on the assumption that the similarity between objects in the network should be reflected in the learned feature representations.

Recommended Read:

- Fire, Michael, et al. "Link prediction in social networks using computationally efficient topological features," 2011
- Altshuler, Yaniv, et al. "How many makes a crowd? On the evolution of learning as a factor of community coverage." International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. Springer, Berlin, Heidelberg, 2012.
- Fire, Michael, et al. "Predicting student exam's scores by analyzing social network data." International Conference on Active Media Technology. Springer, Berlin, Heidelberg, 2012
- Fire, Michael, and Yuval Elovici. "Data mining of online genealogy datasets for revealing lifespan patterns in human population." ACM Transactions on Intelligent Systems and Technology (TIST) 6.2 (2015): 28.
- Primož Godec, Graph Embeddings — The Summary
- Arsov, Nino, and Georgina Mirceva. "Network Embedding: An Overview." arXiv preprint arXiv:1911.11726 (2019)