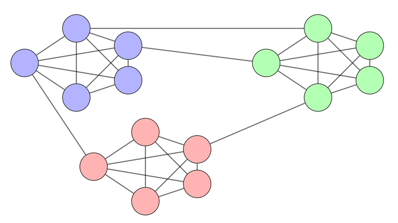
Learning Multiplex Embeddings for Overlapping Community Detection in Online Social Networks

Nouamane Arhachoui[1], Vincent Gauthier[2], Anastasios Giovanidis[1], Lionel Tabourier[1]

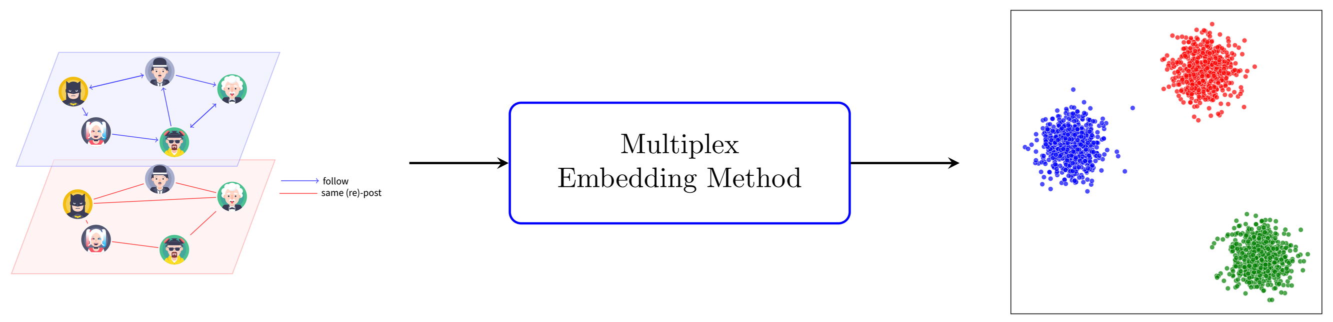
[1]LIP6, Sorbonne Université – CNRS [2]Télécom SudParis, Institut Polytechnique de Paris

Context and Motivation

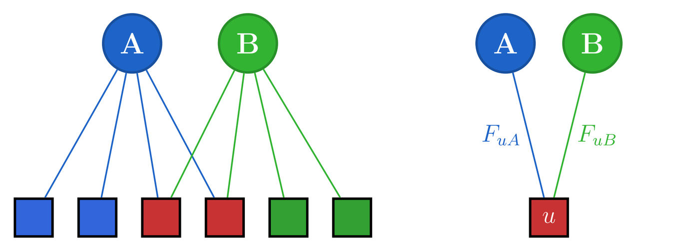
- Community detection identifies groups of nodes that are more densely connected to each other than to the rest of the network



- Most existing methods focus on single-layer networks
- However, Online Social Networks (OSNs) are inherently multiplex, with multiple types of interactions (*e.g.*, follows, reposts, likes, etc.) between the same set of users
- Our goal: leverage the multiplex structure of OSNs to embed users in a latent space that **captures their community memberships across all layers**



Generative Model



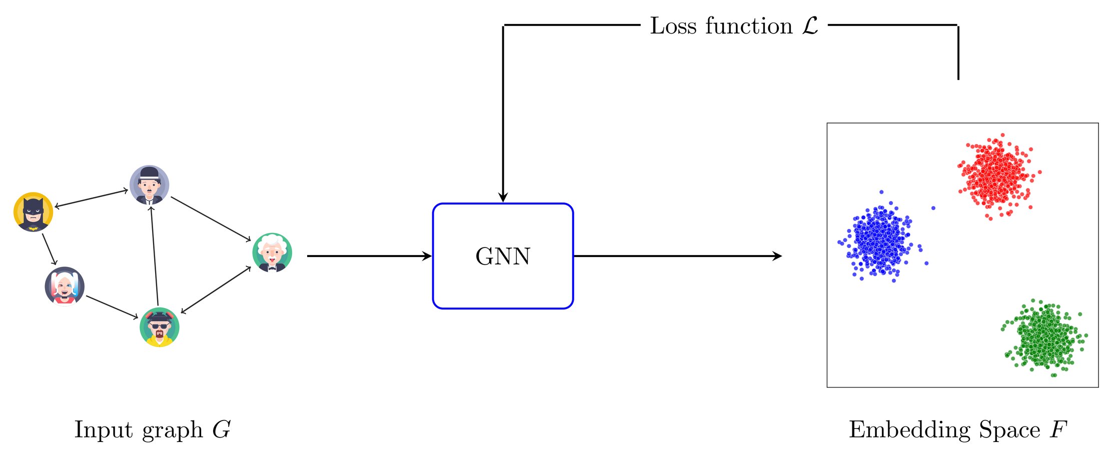
(a) Community Affiliation Network (b) Weight for affiliation

- **BigClam** [3] is a generative model for overlapping community detection in **single-layer networks**. It assumes that each node u has a non-negative affiliation strength F_{uc} to each community c
- The probability of an edge between nodes u and v is given by:

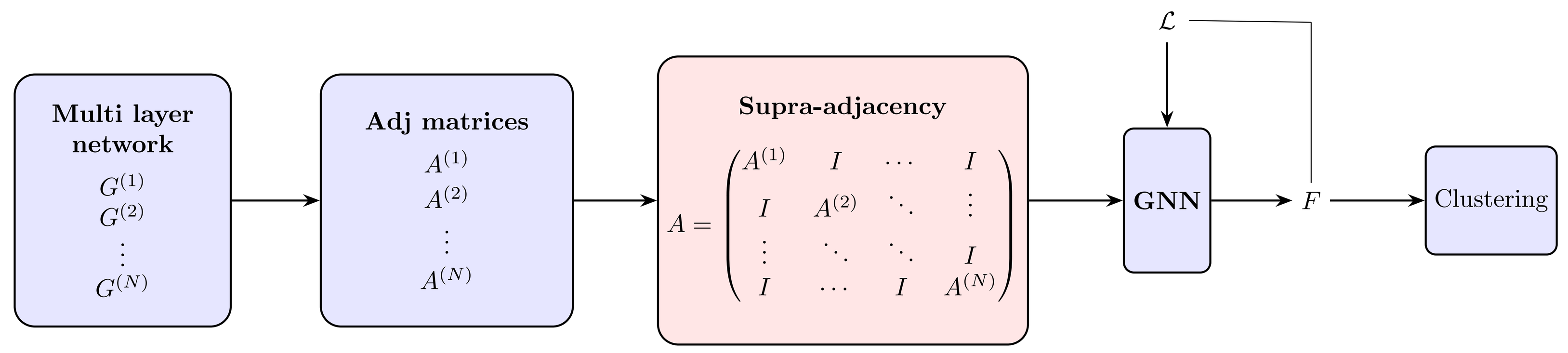
$$P((u,v) \in E | F) = 1 - e^{-F_u^T \cdot F_v}$$

- The model can be trained with a **Graph Neural Network (GNN)** [2] to learn the affiliation matrix F that best explains the observed network structure. The loss function is the negative log-likelihood of the observed edges:

$$\mathcal{L}(F) = -\sum_{(u,v)\in E} \log\left(1 - e^{-F_u^T \cdot F_v}\right) + \sum_{(u,v)\notin E} F_u^T \cdot F_v$$



Proposed Multiplex Extension



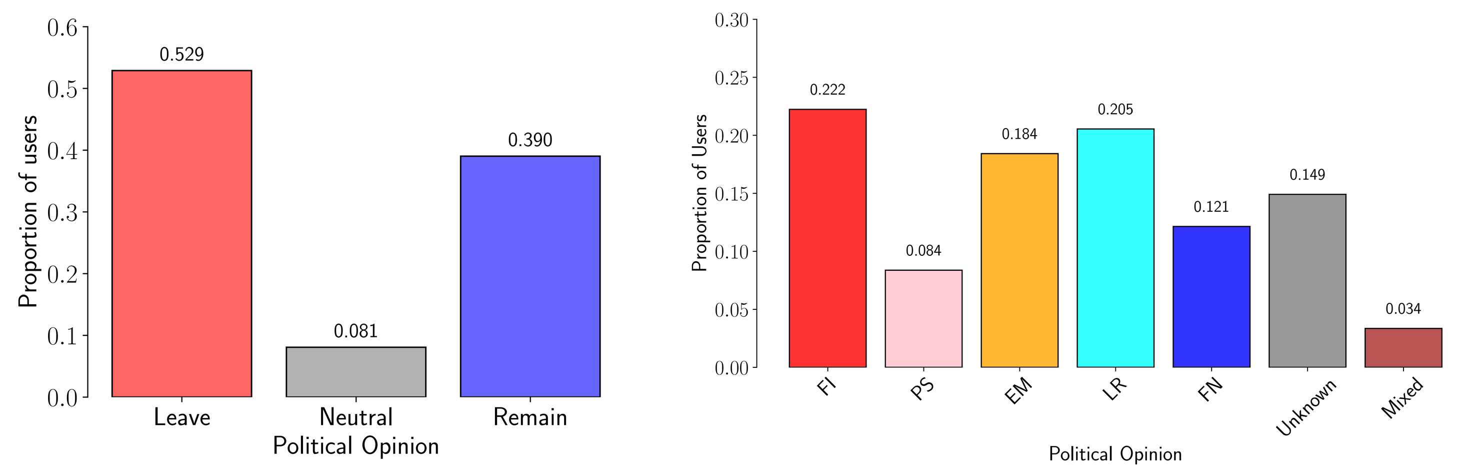
- We build upon the BigClam model by introducing a **multiplex embedding method** that learns a shared affiliation matrix F across all layers of the multiplex network
- The learning process is done by a GNN on the **supra-adjacency matrix** of the multiplex, which jointly captures intra-layer connections and inter-layer identity links, allowing us to leverage the full multiplex structure for community detection.

Real World Applications: Polarization in OSNs

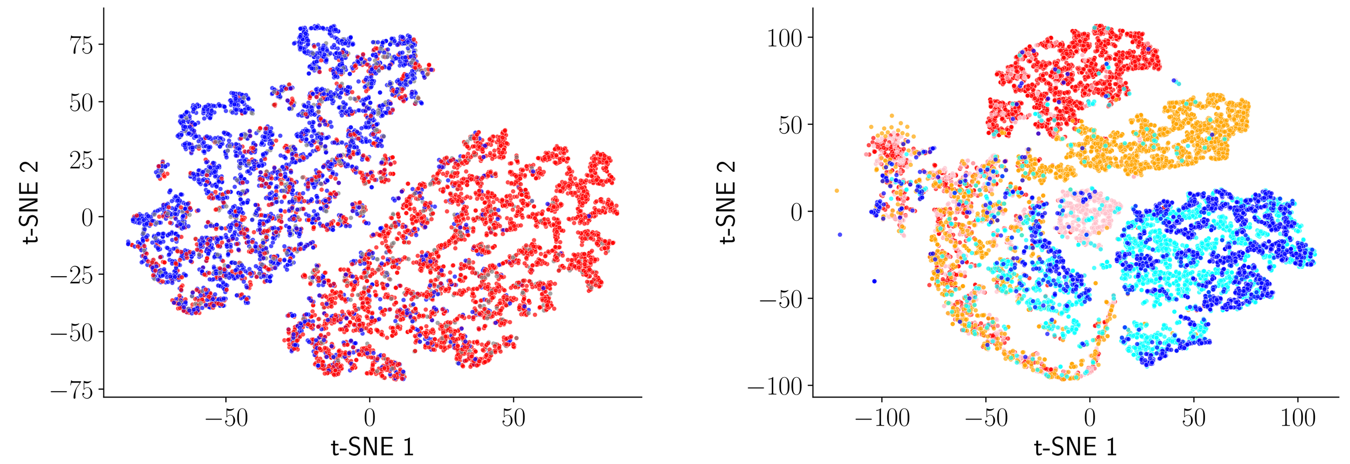
Online discussions on political topics often lead to the formation of polarized communities.

We apply our method to two real-world datasets:

- Brexit, Twitter debate around the UK referendum on leaving the EU $|V| \approx 7.59$k users, $|E_f| \approx 532$k follows, $|E_r| \approx 340$k reposts.
- Elysée, Twitter debate around the 2017 French presidential election $|V| \approx 17.3$k users, $|E_f| \approx 2.41$M follows, $|E_r| \approx 5.33$M reposts.



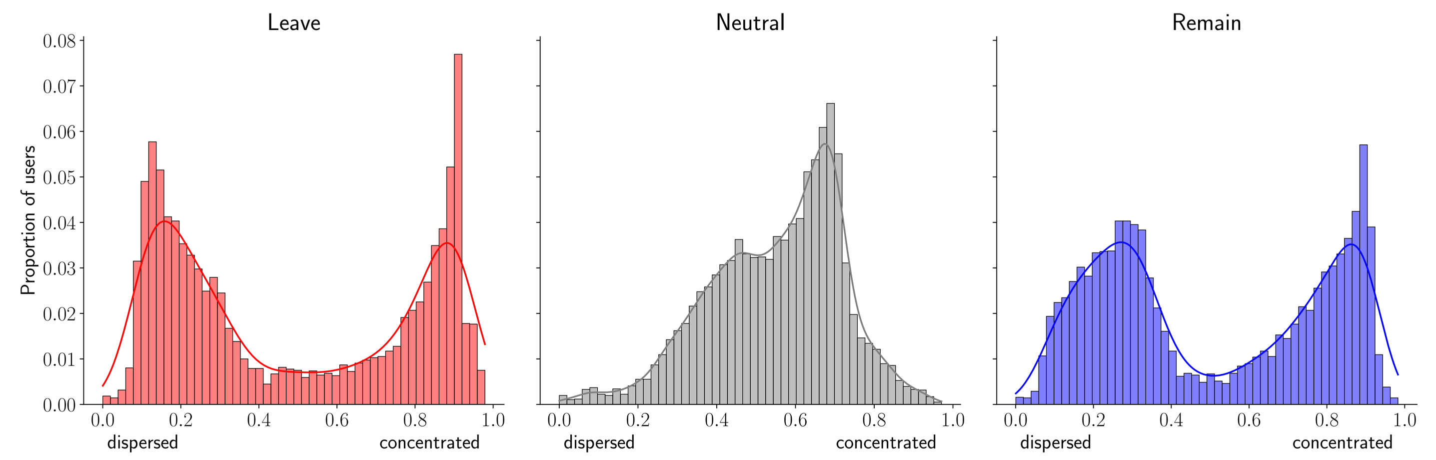
- The **learned embeddings tend to cluster** users with similar political opinions together



- We measure the **reciprocal user diversity** [1] relative to each opinion group centroid, which quantifies how concentrated a user's embedding is around the centroid of a given opinion group:

$$D\left(\tilde{F}_u || \tilde{F}_o\right)^{-1} = \left(\sum_{k=1}^{K} \tilde{F}_{uk}^2 / \tilde{F}_{ok}\right)^{-1}$$

where $\tilde{F}_u$ is the normalized embedding of the user and $\tilde{F}_o$ is the normalized centroid of the opinion group. K is the embedding dimension.



Reciprocal user diversity in Brexit relative to the opinion groups' centroids

- Leave and Remain centroids → clear **bimodal distributions**. Modes are concentrated near 0 and 1, separating users into those concentrated on the same dimensions as the opinion centroid and those who are not. This alignment between embeddings and political opinions suggests that the multiplex structure effectively captures polarization.
- Neutral centroid → the distribution is skewed towards 0.6, suggesting that neutral users do not form a coherent community in the embedding space, but are instead scattered across the polarized communities.

Future Work

This work is part of a broader effort to understand **complex interactions** in OSNs. Our goal is to provide multiplex embeddings that capture these interactions and to further analyze the learned embeddings.
By doing so, we aim to uncover the specific multiplex interaction **patterns that define polarized communities** and distinguish them from neutral or less polarized groups.

References

[1] P. Ramaciotti Morales, R. Lamarche-Perrin, R. Fournier-S'niehotta, R. Poulain, L. Tabourier, and F. Tarissan.
Measuring diversity in heterogeneous information networks.
Theoretical Computer Science, 859:80–115, Mar. 2021.

[2] O. Shchur and S. Günnemann.
Overlapping community detection with graph neural networks.
Deep Learning on Graphs Workshop, KDD, 2019.

[3] J. Yang and J. Leskovec.
Overlapping community detection at scale: a nonnegative matrix factorization approach.
In Proceedings of the sixth ACM international conference on Web search and data mining, WSDM 2013, page 587–596. ACM, Feb. 2013.

nouamane.arhachoui@lip6.fr