# Temporal Analysis of Reddit Networks via Role Embeddings

Siobhán Grayson and Derek Greene

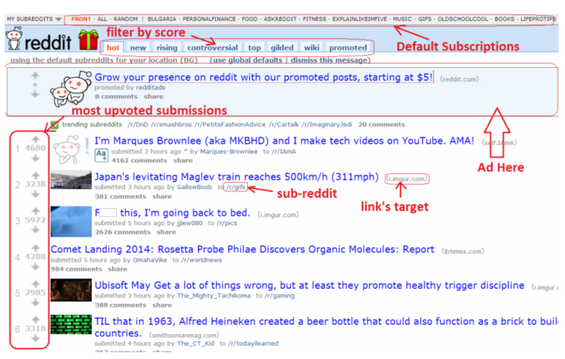
Insight
Centre for Data Analytics

## Motivation

**Inspired by diachronic analysis, we propose an approach for uncovering temporal insights about user roles via graph embeddings.**

reddit

Reddit is a social news aggregation, web content rating, and discussion website.




For this study, we apply the role embedding algorithm, *struc2vec*, to a curated set of subreddits collected over a period of nine months in 2014 exhibiting either "loyal" or "vagrant" characteristics.

We are then able to compare and contrast how user roles change over time by aligning the resulting temporal embeddings spaces. In particular, we analyse temporal role embeddings from an individual and a community-level perspective for both loyal and vagrant communities present on Reddit.

## Methodology

1. Our dataset consists of 16 subreddits identified by Hamilton et al.[1] as exhibiting the most "loyal" user features (teams and sports related subreddits) and 13 subreddits identified as having the highest "vagrant" user patterns.

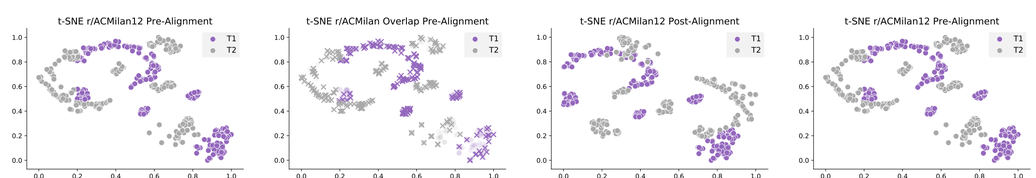
| Class | #SR | #V$_{T1}$ | #E$_{T1}$ | #V$_{T2}$ | #E$_{T2}$ | #V$_{T3}$ | #E$_{T3}$ |
|---|---|---|---|---|---|---|---|
| Loyal | 13 | 15,319 | 89,496 | 15,193 | 91,138 | 14,531 | 87,149 |
| Vagrant | 16 | 13,462 | 22,323 | 14,030 | 23,831 | 13,314 | 22,247 |

Table: Notation - SR: Subreddits, VT 1: Nodes in Temporal Window 1, ET 1: Edges in temporal window 1

2. For temporal analysis, we partitioned 9 consecutive months of data, spanning from late January to October in 2014, into three temporal windows consisting each consisting of three months.

3. Once we have our temporal networks, actors are then described in terms of their roles by applying the directed and weighted version of the graph embedding algorithm, struc2vec [2], specifically designed to capture structural equivalence between nodes.

4. The embedding spaces are then aligned using normalised orthogonal Procrustes which is computed by mapping the overlapping sets of users to each other.



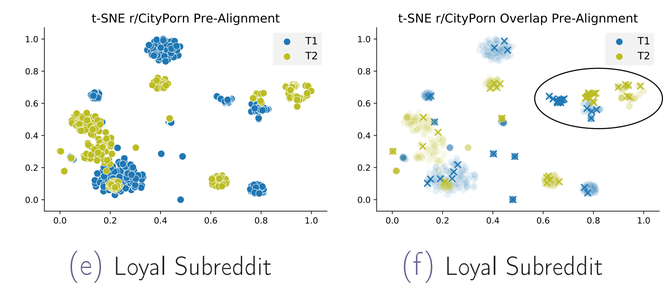

(a) Pre-Alignment (b) User Overlap (c) Post-Alignment (d) PCA

5. To detect changes in an individual's role across time, we compute the cosine distance between an actors embedding across consecutive time windows.

6. Variation of community roles over time is found via analysing how role clusters change across windows using Euclidean Kmeans and kNN.

## Results and Analysis

Although orthogonal Procrustes appeared to perform reasonably well when evaluated using overlapping user embeddings, it did not always correctly align community role embeddings.



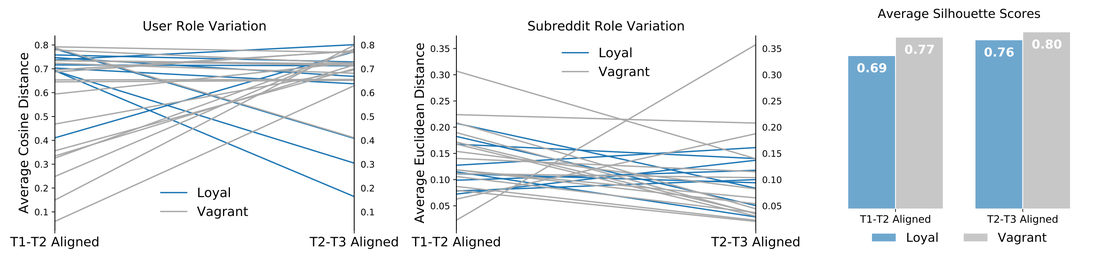

(e) Loyal Subreddit (f) Loyal Subreddit

(g) Vagrant Subreddit (h) Vagrant Subreddit

Unlike ACMilan (Fig. a-d), the majority of overlapping individuals in CityPorn's embeddings are confined to a small region of the space (circled Fig. f).

To resolve this, further alignment of roles is applied by changing the signs of equivalent principal components to agree if they do not already.

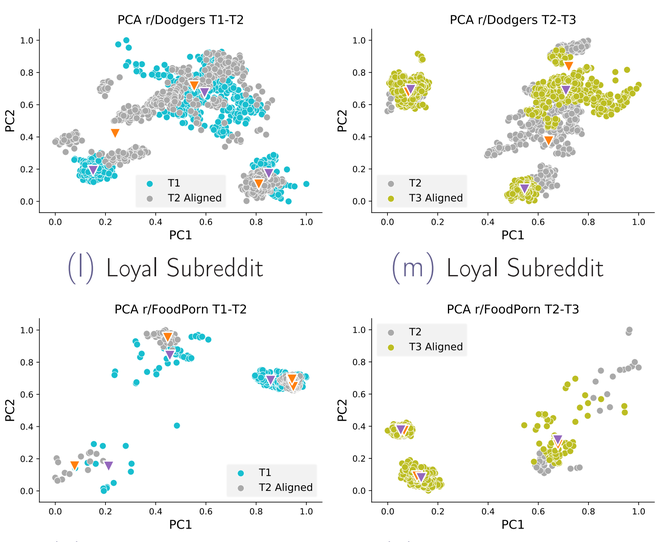
Vagrant users (gray) appear to change roles to a greater extent than loyal users (blue), while loyal users appear to retain the same role over time.




(i) User roles are not static (j) Community roles relatively static (k) Variation of defined roles

The static nature of community roles in comparison to user roles is further examined via average Silhouette Score (Fig. k) with PCA visualisations (Figs. l-o).

The loyal subreddit role clusters are more dispersed in comparison to it's vagrant counterpart where clusters are spread and tightly compact.




(l) Loyal Subreddit (m) Loyal Subreddit

(n) Vagrant Subreddit (o) Vagrant Subreddit

## Conclusion and Future Work

Our preliminary findings suggest that that while participant roles fluctuate a lot, the ubiquitous community roles present are a lot more static. However, further analysis is required and we hope to extend the current work to explore subreddits such as AskReddits, Debate subreddits, Questions subreddits, where distinguished roles may allow for further comparisons to be made.

## References

W. L. Hamilton, J. Zhang, C. Danescu-Niculescu-Mizil, D. Jurafsky, and D. Jurafsky.
Loyalty in online communities.
In *International AAAI Conference on Weblogs and Social Media*, volume 2017, page 540. NIH Public Access, 2017.

L. F. Ribeiro, P. H. Saverese, and D. R. Figueiredo.
Struc2vec: Learning node representations from structural identity.
In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 385–394. ACM Press, August 2017.



**The Insight Centre for Data Analytics is supported by**

**Science Foundation Ireland under Grant Number SFI/12/RC/2289**