Information Diffusion and Temporal Dynamics on Parler

Ashwin Singh

Universitat Pompeu Fabra
Barcelona, Spain
ashwin.singh@upf.edu

1 Introduction

Parler witnessed rapid growth during the 2020 presidential election in United States [1], much of which has been attributed to user migration from popular social media platforms such as Twitter and Facebook [3]. This was not only due to the enforcement of stricter moderation policies to curb the spread of misinformation, but also to deplatform users conspiring to target vote-counting centers and disrupt ballot-tallying [3, 8]. For instance, leaders of notable Facebook groups 'NATIONWIDE RECOUNT 2020' (1.3M users) and 'STOP THE STEAL' (360K users) advocated for migrating to Parler prior to their ban for peddling false election claims in November 2020 [3, 11]. Despite these early signs, this movement materialized into the widely covered Capitol Hill Riots, which aimed to disrupt the electoral college vote-count. [1] In this project, I study two characteristics of the #StopTheSteal movement - (i) information diffusion with a hashtag multigraph and a static user network, and (ii) temporal dynamics of user participation in the movement with a dynamic user network.

2 Sampling Methodology

Parler's user-base grew twice as large during the first week of November 2020, amassing 9M users [1, 3]. This period also coincides with the ban of two major Facebook groups ('NATIONWIDE RECOUNT 2020' and 'STOP THE STEAL'), that cumulatively accounted for over 1.5M users [8], and trending migration hashtags (e.g., 'twexit') due to Twitter's persistent fact-checking of Donald Trump's tweets [3]. Therefore, there is reasonable evidence to suggest the existence of a *migratory* period. Consequently, users who joined Parler after November 2020 are labelled as *migratory* users, and the remaining as *old* users. Top three hashtags (#) with the highest volume in terms of content (posts and comments) on Parler [1] and relevant to the online movement ('stopthesteal', 'electionfraud', 'voterfraud') are used to filter content. Overall, only posts and comments containing these hashtags with creation dates between November 2020 - January 2021 are considered for analysis. More often than not, posts and comments on Parler often contain multiple hashtags. This is leveraged to construct a set of hashtags H and a set of users U from the content containing the top-three hashtags on the platform, resulting in a total of $|H| = 59K$ unique hashtags mentioned by $|U| = 39$K users. However, the top-three hashtags along with hashtags mentioned by exactly one user are both removed to prevent the formation of fully connected and isolated components respectively. More than 50% of hashtags belong in the latter category whereas more than 80% of hashtags are mentioned by < 10 users (cf 1).

3 Network I: Hashtags

Consider a weighted multigraph $G = (H, E)$ where H denotes the set of nodes (hashtags) and E denotes the set of edges. Each edge $e \in E$ is a quadruple (h_i, h_j, w, t) such that hashtags $h_i, h_j \in H$ are mentioned by exactly w unique users in posts or comments. Since edges constitute users, each edge of unit-weight can be either *migratory* or *old* depending on the user. To facilitate a comparative analysis between hashtags connected by *migratory* and *old* users, the multigraph can be partitioned into two graphs. Formally, consider $G_{mig} = (H, E_{mig})$ where an edge $e \in E_{mig}$ is a triple (h_i, h_j, w) such that both hashtags $h_i, h_j \in H$ are mentioned by exactly w unique *migratory* users in posts or comments before time t. Similarly, consider $G_{old} = (H, E \setminus E_{mig})$.

[1] The Electoral College vote-count is the final step to confirm the President-elect in the United States.

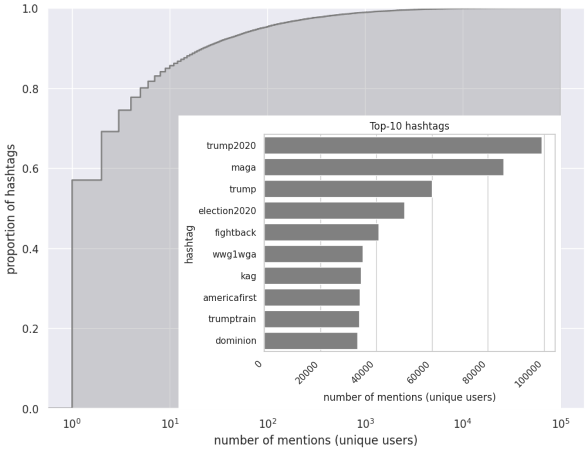

Fig. 1: CDF plot for number of hashtag mentions by users with an inset bar plot of top-10 most popular hashtags. 80% hashtags are mentioned by < 10 users.

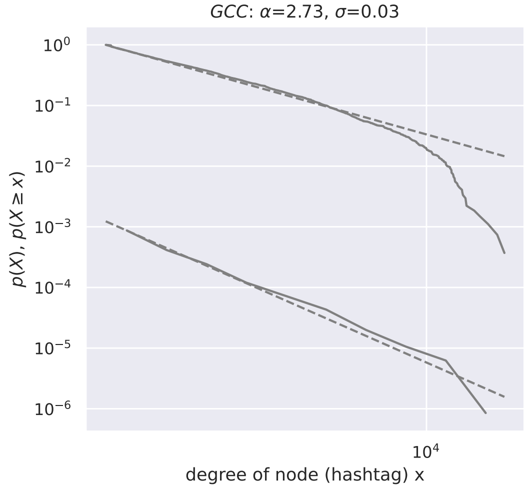

Fig. 2: CCDF and PDF plots for the Power law fit of the degree distribution of GCC

Only the giant connected components, namely GCC, GCC_{mig} and GCC_{old} extracted from G,G_{mig} and G_{old} respectively are considered for analysis. GCC_{mig} and GCC_{old} share over 10K nodes (hashtags) in common, and are naturally subgraphs of GCC. The degree distribution of all three giant connected components follows a power law (cf 2 for the power law fit for GCC). GCC_{mig} has the lowest value for $\alpha \approx 2.69$ (lowest skew), or the least disparity in terms of co-mentions or degrees of hashtags. For the ease of analysis, edges which constitute less than ten users are filtered out and giant connected components are recomputed until all three components share the same set of nodes (hashtags). Essentially, the relationship between any two hashtags is considered significant only if they are co-mentioned by at least ten users.

3.1 General Characteristics

The general characteristics of GCC, GCC_{mig} and GCC_{old} before and after filtering are reported in Table 1. The analysis hereon focuses on the recomputed components following the removal of insignificant edges ($w < 10$). Table 2 lists the most central nodes in the three giant connected components and distance measures in the three networks.

Table 1: General characteristics for GCC, GCC_{mig} and GCC_{old} before and after filtering.

	E **[unfiltered]**			$E' = \{e = (h_i, h_j, w) \in E \mid w \geq 10\}$		
	# of nodes	# of edges	density	# of nodes	# of edges	density
GCC	18.2K	6.5M	3.9%	1216	149.7K	20.2%
GCC_{mig}	12.6K	2.6M	3.2%	1216	64.8K	8.7%
GCC_{old}	15.6K	5M	4%	1216	84.7K	11.4%

First, weighted centrality measures are used to compute the most central nodes in all three networks:

1. **Closeness Centrality:** With the highest closeness centrality in GCC, *#ilovemypresident* reflects the common sentiment across both *migratory* and *old users* on Parler in the favor of Donald Trump. On the other hand, hashtags *#Montana* (US state) and *#AntrimCounty* in Michigan were both subject to claims of voting irregularities and fraud during the Presidential Election [9, 7]. As nodes, these hashtags are strategically best placed to broadcast information due to their high closeness centrality. This shows that *migratory* and *old users* focused on targeting different states with conspiracy theories.

2. **Betweenness Centrality**: *#WakeUpAmerica* in GCC and *#AmericaFirst* in GCC_{old} reflect strong nationalist sentiments that bridge different hashtag topics in both networks. *#Dominion* in GCC_{mig}

Table 2: Most central nodes and other statistical properties of GCC, GCC_{mig} and GCC_{old}.

		GCC	GCC_{mig}	GCC_{old}
Most Central Nodes	**Closeness**	#ilovemypresident	#montana	#antrimcounty
	Betweenness	#wakeupamerica	#dominion	#americafirst
	Pagerank	#dominion	#maga2020	#dominion
Average Path Length		1.79	1.92	1.89
Diameter		2	4	4
Effective Diameter		2	2	2

refers to the American company responsible for producing electronic voting machines for the elections.

3. **Eigenvector Centrality (Pagerank)**: It is interesting to note that *#Dominion*, which has the highest betweenness centrality in GCC_{mig}, has the highest influence over the entire network in both GCC and GCC_{old}. Lastly, *#MAGA2020* is another popular hashtag among Republican supporters which was first used by Ronald Reagan during his presidential campaign in 1980.

Second, distance measures are used for the comparison of average path lengths and diameters of the three networks:

1. **Average path length**: While GCC naturally has the shortest average path length, it is interesting to note that the same set of hashtags that represent a coordinated campaign are better connected by *old* users in comparison with *migratory* users.

2. **Diameter**: While all three connected components share the same effective diameter (90%) of two, both GCC_{mig} and GCC_{old} have twice the actual diameter of GCC, highlighting the importance of coordinated posting on Parler between *migratory* and *old* users.

3.2 Information Flows

Since GCC_{mig} and GCC_{old} share the same set of nodes H, comparing their edge sets E_{mig} and $E \setminus E_{mig}$ can help understand how *migratory* and *old* users differ in connecting information concerning any two hashtags in GCC. For the same, Jaccard Similarity between neighborhoods of each hashtag in GCC_{mig} and GCC_{old} is computed. More specifically, if the neighborhood of a hashtag $h \in H$ has node set $\Gamma_{mig}(h)$ in GCC_{mig} and $\Gamma_{old}(h)$ in GCC_{old}, then Jaccard Similarity (s) of h is defined as follows:

$$s(h) = \frac{|\Gamma_{mig}(h) \cap \Gamma_{old}(h)|}{|\Gamma_{mig}(h) \cup \Gamma_{old}(h)|} \tag{1}$$

Hashtags that exhibit high Jaccard Similarity (s $>$ 0.8) i.e., hashtags whose information is similarly connected to other hashtags by *migratory* and *old* users are highlighted in the CCDF plot of s (cf 3). At a closer look, these are broadly concerned with four themes - **support for Trump, conspiracy theories, nationalist propaganda** and content created by **conservative political commentators**. Next, to identify important hashtags whose information is connected differently by *migratory* and *old* users to other hashtags, these high similarity hashtags (s $>$ 0.8) are excluded from the node set H. Since Gephi does not currently support multigraphs, a logistic function (σ) is applied to the difference between edge-weights of the two edgetypes in GCC to obtain a new edge-weight as shown below:

$$\sigma(w_{migratory} - w_{old}) = \frac{1}{1 + e^{-(w_{migratory} - w_{old})}} \tag{2}$$

Intuitively, this represents how much an edge between any two hashtags (nodes) in GCC leans towards *migratory* or *old* users. Evident from the distribution of σ in Figure 4, *migratory* users only dominate about 30% edges in GCC. As a corollary, despite being similar in number, migratory users are only

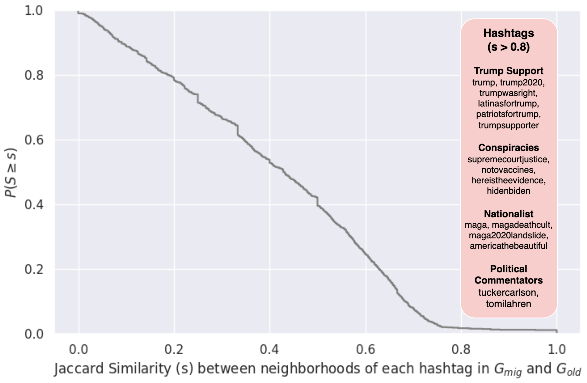

Fig. 3: A CCDF plot of Jaccard Similarity (s) between neighborhoods of each hashtag in GCC_{mig} and GCC_{old}. Hashtags with high similarity (s > 0.8) are thematically grouped.

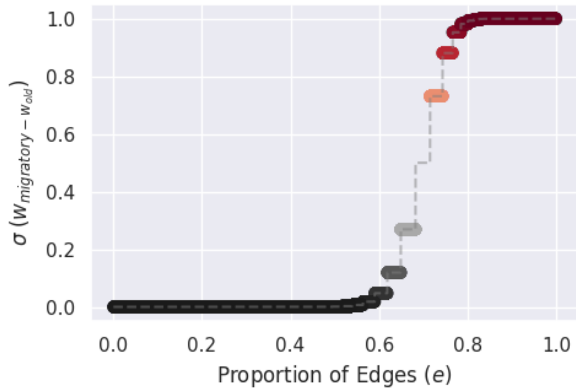

Fig. 4: Distribution of the logistic function (σ) applied to the difference between *migratory* and *old* edge-weights. Redness denotes the intensity of the information flow in an edge driven by *migratory* users.

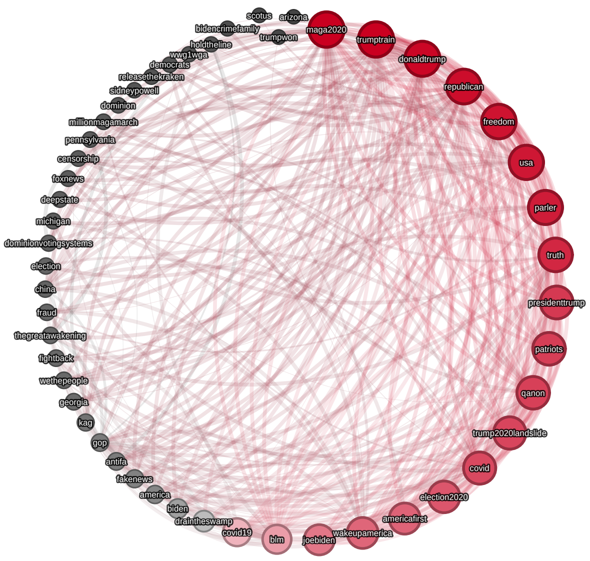

Fig. 5: Top-50 hashtags where *migratory* users dominate information flow visualized on a **circular layout** using Gephi. Edge weights are scaled based on the logistic function σ (equation 2) whereas the node sizes and redness are scaled based on the resulting weighted degree.

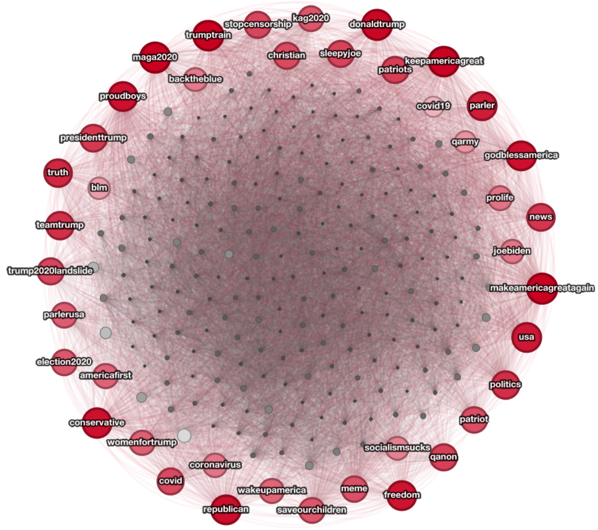

Fig. 6: k-core ($k = 193$) of hashtag network containing 247 nodes (hashtags). Hashtags where old users dominate information flow form the majority in the k-core.

responsible for about 30% of information flow in the network of hashtags concerning the online movement on Parler. The hashtags most dominated by *migratory* users in terms of information flow (weighted degree based on σ) are represented in Figure 5. Note that it follows the same gradient as Figure 4, where both the redness of color and size of node are proportional to the weighted degree of the hashtag. Finally, the k-core of the network is visualized (cf 6) in Gephi using Force-Atlas 3D and circular layout respectively. In the former, there exists a k-core with value 193 and size 247. Applying the same logistic coloring from Figures 4 and 5, it appears that hashtags where *old* users dominate information flow form the majority in the k-core. This implies hashtags most efficiently placed in the global topology of the network in terms of passing information rely predominantly on *old* users. Node size is scaled based on higher weighted degree (more *migratory* users driving flow) to highlight the remaining hashtags.

4 Network II: Parler Users

In this section, the remaining 1060 hashtags where *migratory* and *old* users drive information flow differently (s $<$ 0.8 from equation 1) are utilized to study networks at the user-level. Consider a weighted undirected network $G = (U, E)$ where U denotes the set of users who mention at least ten but no more than 50 out of the 1060 hashtags; $|U| = 6602$. This is to both filter out users with negligible contribution as well as users whose posting behavior mimics spamming. Each edge $e \in E$ is a triple (u_i, u_j, w) such that users u_i, u_j mention $w \geq 10$ common hashtags in their posts. Essentially, the relationships between any two users is considered significant only if they mention at least ten common hashtags in their posts. Upon extracting the giant connected component satisfying these conditions, there are 1605 users (725 *migratory* and 880 *old*) left with 24.8K edges between them (density = 1.9%). The giant component has an average path-length $\approx$ 2.81 and diameter=8. Interestingly the former is nearly the same as the average path-length between *migratory* and *old* users in the network. In the giant component, nearly 80% users have degree $\leq$ 60 (cf 7 for the log-log plot of degree distribution where the dashed vertical line represents degree=60). To get a better topological view of the network, highly connected nodes (degree $>$ 60) are excluded and the ForceAtlas 3D layout in Gephi is utilized (cf 8). This eliminates about 90% edges in the network, and results in several isolated components, highlighting their importance in the network.

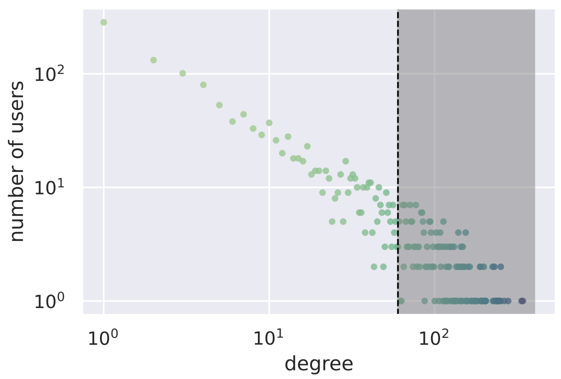

Fig. 7: Log-log plot for the degree distribution of Parler users. 82.7% users have degree $\leq$ 60 - denoted by the vertical dashed line in the plot.

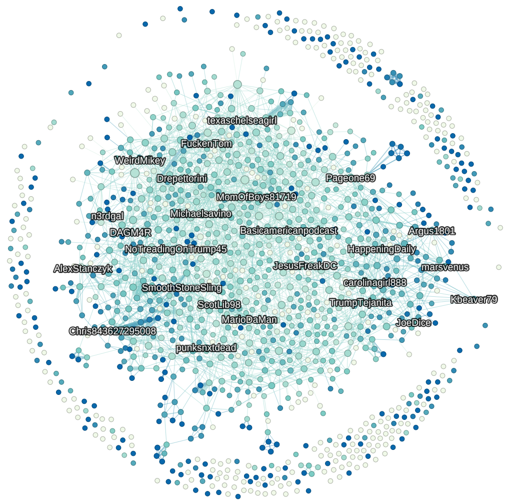

Fig. 8: Parler users network visualized using a ForceAtlas 3D layout on Gephi after excluding highly connected nodes (degree $>$ 60).

4.1 Communities

The extracted giant component is decently modular ($\approx$ 0.55 on average across ten runs), resulting in three major communities that capture $\approx$ 98% of the nodes in the network (cf 9). Node sizes are color-coded by modularity class and scaled in size based on eigenvector centrality. On the right, bar plots of the top five hashtags corresponding to the three communities are visualized to better contextualize the topics within them. Community I focuses on the hunter biden laptop controversy (#hunterslaptop), which was published in The New York Post three weeks before the Presidential Election [4]. The story was an attempt by Trump and his personal attorney Rudy Giuliani to hurt Joe Biden's campaign using a corruption ploy related to Ukraine and his son Hunter Biden. [2] Community II is ampliying the comments made by Sidney Powell (#krakenonsteroids) who was indicted along with Donald Trump for conspiring to unlawfully change the outcome of the election in Georgia through criminal racketeering [6]. Lastly, Community III has users supporting Charlie Kirk, who was endorsed by Trump's attorney Rudy Giuliani for leading #stopthesteal rallies. His organization 'Turning Point USA' was responsible for sending several buses of rioters to the Capitol Hill on January 6th [10].

[2] https://en.wikipedia.org/wiki/Hunter_Biden_laptop_controversy

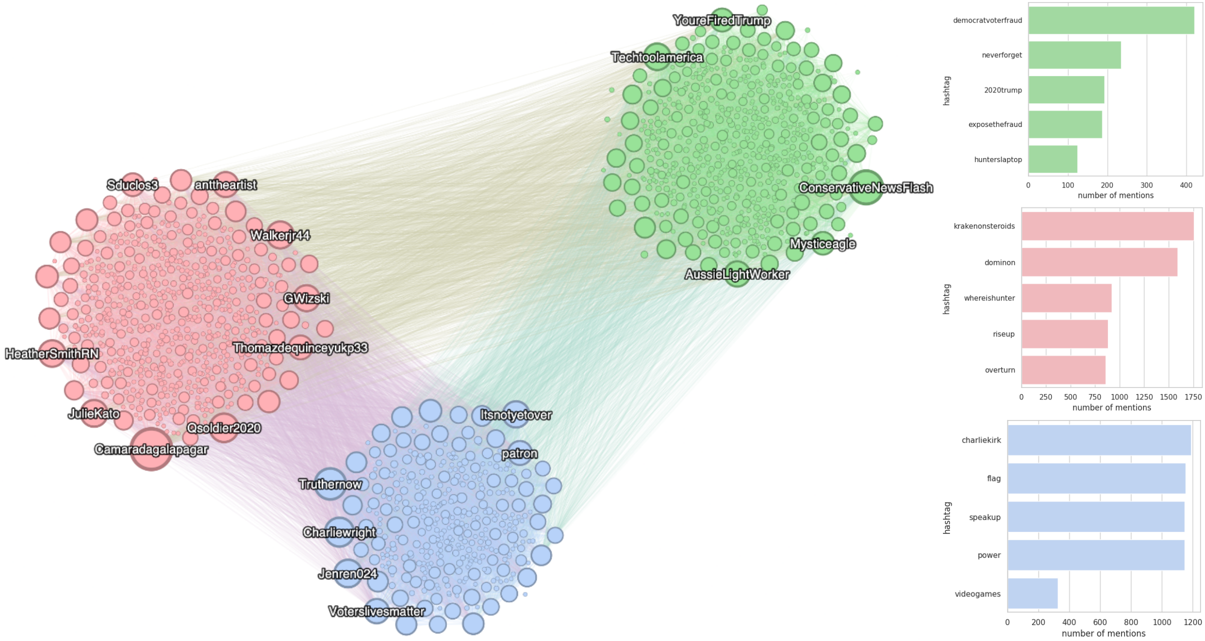

Fig. 9: Node size ∝ **eigenvector centrality**, where the top-20 nodes with the highest centrality are labelled. Top five hashtags characterizing each community are visualized in a bar plot to the right.

4.2 Assortativity and User Roles

What roles do *migratory* and *old* users play in the network based on centrality measures? Are the differences in their respective behaviors significant? What is their tendency to form edges among themselves or each-other? This section attempts to explore and answer these questions.

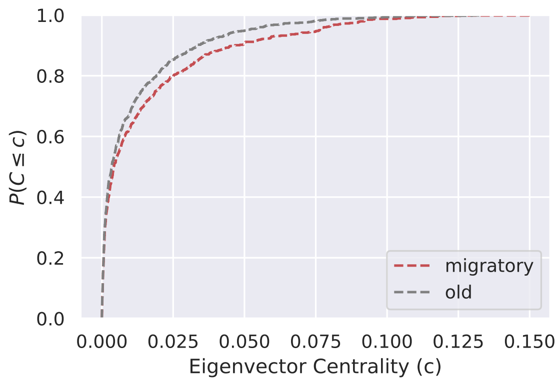

Fig. 10: Empirical CDF plot of eigenvector centralities for both *migratory* and *old* users in the network. Certain *migratory* users exhibit higher influence.

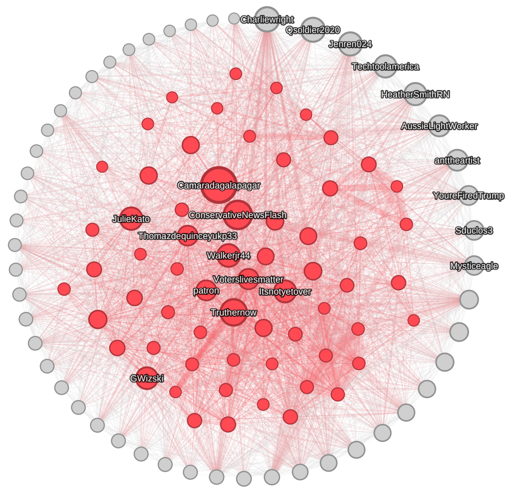

Fig. 11: Node size ∝ **eigenvector centrality**, where nodes with high eigenvector centrality are labelled. With an assortativity coefficient ≈ 0.045, there is near neutral mixing between *migratory* (red) and *old* users (gray).

First, a two-sample Kolmogorov-Smirnov test is conducted to understand *migratory* and *old* exhibit different centrality distributions. In the case of closeness and betweenness centrality, the test finds no significant differences ($p > 0.05$). However, in the case of eigenvector centrality, there is sufficient evidence to rule out the null hypothesis that *migratory* users exhibit less influence than *old* users in the network. Essentially, the trend that certain *migratory* users exhibit higher influence is statistically significant ($p < 0.05$), which can be observed in their CDF (cf 10). Second, the assortativity coefficient for the property whether the user is *migratory* or *old* is ≈ 0.045, suggesting no clear preference for forming edges or a nearly neutral mixing between the two types of users. These connections are visualized in Figure 11, where *migratory* users are colored red, whereas *old* users are colored gray. Node sizes are scaled by eigenvector centrality; *migratory* users are organized in a Fruchterman Reingold layout within the circular layout formed by *old users*.

5 Temporal User Network

In this section, timestamps are leveraged from posts and comments to unpack the temporal dynamics of the Parler user network. For this exercise, all available data i.e., 337K posts and comments made by 39K users, is used without any filtering. [3] First, all 'edge-forming' hashtags i.e., hashtags mentioned by more than one unique user are extracted from the posts for each time-step (day); $|H| = 17K$ out of 59.2K hashtags. Second, all unique users who mention at least one of the 'edge-forming' hashtags in the respective time-step (day) are obtained. Using this streaming approach, the number of active users i.e., users who form edges based on common hashtags are captured for each time step (day). Similarly, 'active' hashtags (present in edges between users) are also captured for each day from November 1, 2020 until January 9, 2021. Figure 12 shows the percentage of daily active users and hashtags within this time frame. Both follow a similar trend, peaking in the second week of November and slowly decaying until $< 15\%$ hashtags and $< 5\%$ users are active. Activity levels of both users and hashtags are lowest following a sudden dip during Christmas period.

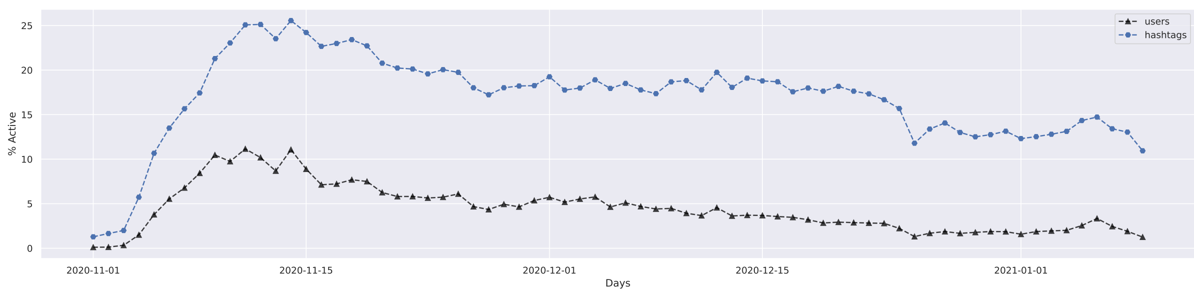

Fig. 12: Time series plot showing percentage of active users and hashtags from November 1, 2020 until January 9, 2021. Both follow a similar trend, with the highest activity levels on November 14, 2020 and the lowest activity levels on December 25, 2020.

Next, the users are again partitioned into *migratory* and *old* users to understand the differences, if any, in their hashtag mentioning behavior through time. Interestingly, the *old* users are almost always more active than *migratory* users until the first week of December, when their activity levels are indistinguishable from each other till the Capitol Hill Riots (cf 13).

6 Modelling Information Diffusion

In this section, the information spread in the temporal user network is studied using the SIR (Susceptible-Infected-Recovered) model. More specifically, if U_t denotes the set of users who form an edge in the network at time t, the transmission rate ν, and recovery rate δ are defined as follows:

$$\nu(t) = \frac{|U_t \setminus U_{t-1}|}{|U_{t-1}|} \quad \text{and} \quad \delta(t) = \frac{|U_{t-1} \setminus U_t|}{|U_t|} \tag{3}$$

[3] This 'all available data' refers to the initially sampled data (§ 2) using hashtags relevant to the #StopTheSteal movement on Parler.

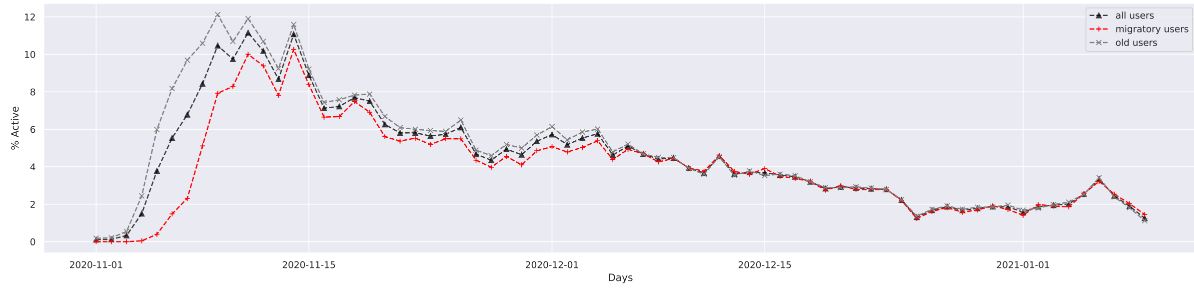

Fig. 13: Time series plot showing percentage of active users - *migratory*, *old* and overall from November 1, 2020 until January 9, 2021. *Old* users are more active than *migratory* users until the first week of December, when their activity levels become indistinguishable till the insurrection.

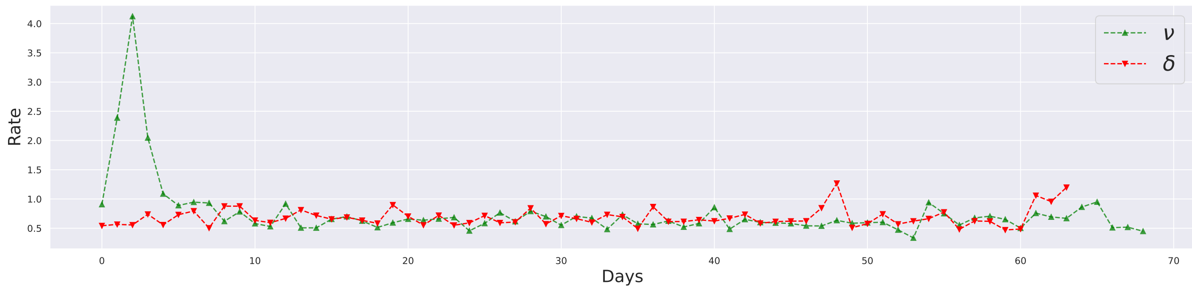

Fig. 14: Time series plot of transmission rate ν and recovery rate δ.

Essentially, a user is considered infected if it forms at least one edge at time t, and recovered if it forms no edges at time t. As current implementations of SIR are limited by static growth and recovery rates, the average of the empirically observed rates through time is used to estimate these parameters. However, the user network is afforded $t = 7$ days or a one-week period for initial growth. This is because of two reasons. First, it addresses the cold start problem (cf 12), and allows the network to contain a considerable number of 'initially infected nodes' (ρ). Second, it bypasses the anomalous growth rate observed in the first week of November (cf 14). For the remaining timesteps (63 days), the overall transmission rate $\mathcal{V}$, and recovery rate $\varDelta$ are approximated using the averages of ν and δ as follows:

$$\mathcal{V} = \frac{1}{T-7}\sum_{t>7}^{70}\nu(t) \approx 0.642 \quad \text{and} \quad \varDelta = \frac{1}{T-7}\sum_{t>7}^{70}\delta(t) \approx 0.682 \tag{4}$$

Finally, in the equation for SIR model: $S(t) + I(t) + R(t) = N$, the following initial conditions are used:

$$I(t=0) = \rho = 0.0678N\ ^{4} \quad \text{and} \quad R(t=0) = 0 \tag{5}$$

6.1 Network Generation

In this section, the followers of Parler users are leveraged as a degree sequence to generate a network using the configuration model. This helps address the computational challenge of constructing the network with over 39K nodes (users). First, a random sample of 10K users is extracted, and a Kolmogorov-Smirnov test is used to validate its goodness of fit with the empirically observed degree sequence ($p > 0.99$ cf 15). Second, the configuration model is used to generate a network with this degree sequence. The giant component in this generated network has 9.83K nodes (users), and 619K edges, with an average path length of 2.04 and diameter of 4 ($D_{eff} = 2$). These characteristics are somewhat akin to a small world phenomenon in the network.

[4] At time t=7, 6.78% users are infected -cf 12.

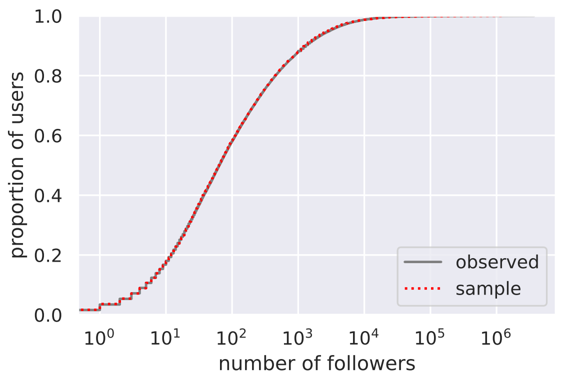

Fig. 15: CDF plot of the random sample and the observed follower distribution in Parler data.

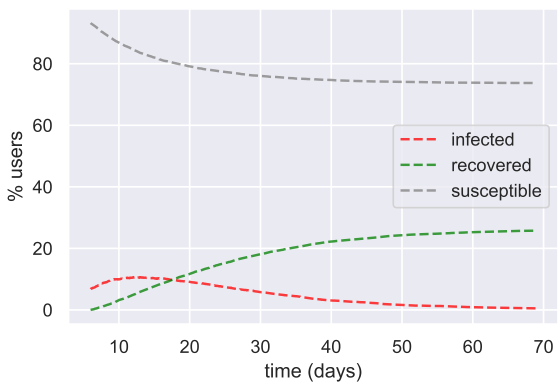

Fig. 16: Time series plot of $S(t), I(t), R(t)$ generated using the best fit of SIR on the empirical data.

6.2 Grid Search for Parameter Learning

While parameters for simulating SIR are available from from equation 4, they are generated using a network where edges are formed between users based on common hashtags and not following behavior. In the real-world, only a fraction of a user's followers may mention hashtags in their posts or participate in the online movement. In fact, the percentage of posts containing hashtags in the entire dataset is less than 3% [1]. Ideally, both $\mathcal{V}$ and Δ should be discounted by computing the proportion of a user's followers who post or don't post hashtags at each time-step. However, that is computationally infeasible given the scale of the Parler dataset. Therefore, grid search is performed on top of SIR, using a logarithmic range of discounts - α for $\mathcal{V}$ and β for Δ respectively. The best fit on the empirical data (cf 12) is determined using a simple least-squares error function, yielding $\alpha = 0.0005$ and $\beta = 0.125$. The corresponding time series for the susceptible $S(t)$, infected $I(t)$ and recovered $R(t)$ users can be seen in Figure 16.

To further validate the best-fit, it is plotted with the empirical data (cf 17) for two-cases of initially infected users: **random** and **highest degree**. Evident from the plot, initially infecting the **highest degree** nodes better captures the temporal trend of information diffusion. This also implies that the SIR model is highly sensitive to the initially infected nodes in networks with a heavy tailed degree distribution. In context of Parler, it means that the information diffusion for any online movement heavily relies on the choice of initial nodes. Choosing the **highest degree** users can result in an overall higher peak of participating (infected) users.

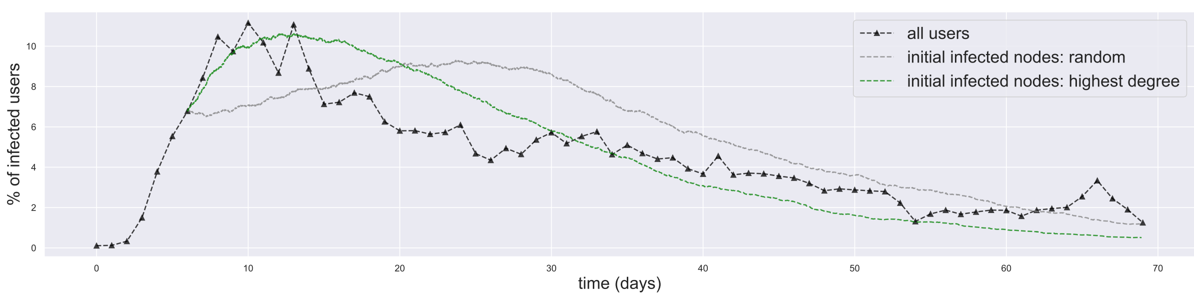

Fig. 17: Time series of the best fit of SIR on the empirical data for two-cases of initially infected nodes. Initially infecting the highest degree nodes better captures the temporal trend of information diffusion. Initially infecting the highest degree nodes better captures the temporal activity of users.

6.3 Understanding Parameters

The learned parameters α, β in § 6.2 or discounts for $\mathcal{V}, \Delta$ are a correction for the rates based on the proportion of a user's participant / non-participant followers in the online movement for each time-step. More specifically, they capture the following:

$$\alpha = \sum_t \sum_{u \in U} \frac{\mid \Gamma_t(u) \setminus \Gamma_{t-1}(u) \mid}{|deg(u)|} \quad \text{and} \quad \beta = \sum_t \sum_{u \in U} \frac{\mid \Gamma_{t-1}(u) \setminus \Gamma_t(u) \mid}{|deg(u)|} \tag{6}$$

Note that here, Γ_t refers to the neighborhood of user u in the network where edges are represented by common hashtags at time t, whereas $deg(u)$ is the degree of user u in the follower network generated using the configuration model. The limitation of this modelling is that it assumes that a user can only infect their followers through hashtags and does not take into account, the content ranking algorithm of Parler. However, with just two additional parameters, the SIR model was able to reasonably capture temporally, the information diffusion in the follower network.

6.4 SIR Model Animation

The learned parameters in § 6.2 are scaled to produce a comparable SIR timeseries curve, similar to Figure 16 for a 2D grid graph of 10K nodes. Note that this is only for the purpose of visualizing the $S(t), I(t), R(t)$ through time and not otherwise representative of information diffusion in the real network. Due to the nature of the deliverable it is not included in this report but instead included in the presentation which can be found here on slide 19.

7 Discussion

In this project, I analyzed data from the Parler network for a period of 70 days (Nov 1, 2020 - January 9, 2021) until the #StopTheSteal movement materialized into the Capitol Riots. In particular I constructed and analyzed three networks, whose findings are summarized below:

1. **Hashtag Multigraph**: I found that *migratory* and *old* users targeted different states for broadcasting conspiracy theories about voter fraud (§ 3). However, despite being similar in number, *migratory* users only dominate information flow for about 30% edges in the hashtag network (§ 3.2). At the same time, hashtags where old users dominate information flow form the majority in the k-core of the network.

2. **User Network**: I found that *migratory* and *old* users do not have a strong preference for posting the same hashtags as each other (§ 4.2). However, *migratory* users exhibit higher influence on the entire network. Lastly, communities of users focus on propagating different conspiracy theories through the network (§ 4.1).

3. **Temporal User Network**: I studied the temporal activity of both users and hashtags in the dynamic user network using a streaming approach (§ 5). The information diffusion in this temporal network was successfully modelled using a two-step approach. First, the degree sequence based on user followers was leveraged to produce a network using the configuration model (§ 6.1). Second, by adding two additional parameters in grid search on top of the SIR model, the information diffusion was successfully simulated with a reasonable accuracy (§ 6.2). Finally I found that the SIR model is highly sensitive to the choice of initial nodes in a network with a heavy tailed degree distribution. This means that the information diffusion of any topic or online movement heavily relies on the choice of initial users on Parler.

Nevertheless, there are some limitations to this work. For instance, the assumption that a user can only infect their followers does not take into account the content ranking algorithm or external factors. Similarly, this work does not account for the differential velocity of information spread across hashtags i.e., hashtags posted by more users may spread faster in the user network. However, this is also a limitation of the current implementations of the SIR model as they lack support for different transmission rates for edges in the network. At the same time, it is also computationally difficult to construct a network of that order. I acknowledge these limitations and that it was difficult to address them given the scope of this course project.

Bibliography

[1] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. A large open dataset from the parler social network. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):943–951, May 2021.

[2] donk_enby. Parler tricks. https://doi.org/10.5281/zenodo.4426283, January 2021.

[3] Elizabeth Dwoskin and Rachel Lerman. 'stop the steal' supporters, restrained by facebook, turn to parler to peddle false election claims. *Washington Post*, 2020.

[4] Adam Goldman. What we know and don't about hunter biden and a laptop. *The New York Times*, 2020.

[5] Jay Greene, Rachel Lerman, and Tony Romm. Seldom-seen amazon unit made the call that brought down pro-trump parler. *Washington Post*, Jan 2021.

[6] Hugo Lowell. Trump's ex-lawyer sidney powell pleads guilty in georgia election case. *The Guardian*, 2023.

[7] Craig Mauger. Michigan expert debunks infamous report on antrim county election as 'meaningless'. *The Detroit News*, 2021.

[8] Stanley-Becker Isaac Romm, Tony and Elizabeth Dwoskin. Facebook bans 'stop the steal' group trump allies were using to organize protests against vote counting. *Washington Post*, 2020.

[9] Alex Sakariassen. Voter fraud: Fact, or faction? *Montana Free Press*, 2021.

[10] Jospeh Tanfani. How trump's pied pipers rallied a faithful mob to the capitol. *Reuters*, 2021.

[11] Brandy Zadrozny. Misinformation by a thousand cuts: Varied rigged election claims circulate. *NBC News*, Nov 2020.