Chapter 1

The growth of social groups

1.1 Social groups

Two popular online platforms **Reddit** and **Meetup** are organized into different groups. On Reddit [1], users create subreddits, where they share web content and discussion on specific topics, so their interactions are online through posts and comments. The Meetup groups [2], are also topic-focused, but the primary purpose of these groups is to help users in organizing offline meetings. As meetings happen face-to-face, Meetup groups are geographically localized, so we'll focus on groups created in two towns, London and New York.

The Meetup data cover groups created from 2003, when the Meetup site was founded, until 2018, when using the Meetup API we downloaded data. We extracted the groups from London and New York that were active for at least two months. There were 4673 groups with 831685 members in London and 4752 groups with 1059632 members in New York. For each group, we got information about organized meetings and users who attended them. From there, for each user, we can find the date when the user participated in a group event for the first time; it is considered the date when the user joined a group.

The Reddit data were downloaded from https://pushshift.io/ site. This site collects posts and comments daily; data are publicly available in JSON files for each month. The selected subreddits were created between 2006 and 2011, we also filtered those active in 2017. We removed subreddits active for less than two months. The obtained dataset has 17073 subreddits with 2195677 active members. For each post, we extracted the subreddit-id, user-id and the date when the user created the post. Finally, we selected the date when each user posted on each subreddit for the first time.

1.1.1 The empirical analysis of social groups

For each Meetup group we have information when user attended the group event, while for subreddit we have detailed data about user activity, so we can extract the information when user for the first time created a post. Those dates are considered as timestamp when user joined to group. So both datasets have the same structure: (g, u, t), where t is timestamp when user u joined group g. For each

[1] https://www.reddit.com/
[2] www.meetup.com

time step, we can calculate the number of new members in each group $N_i(t)$, and the group size $S_i(t)$. The group size at time step t is $S_i(t) = \sum_{k=t_0}^{k=t} N_i(t)$, where t_0 is month when group is created. The group size is increasing in time, as we do not have information if the user stopped to be active. Also we calculate the growth rate, as the logarithm of successive sizes $R = log(S_i(t)/S_i(t-1))$.

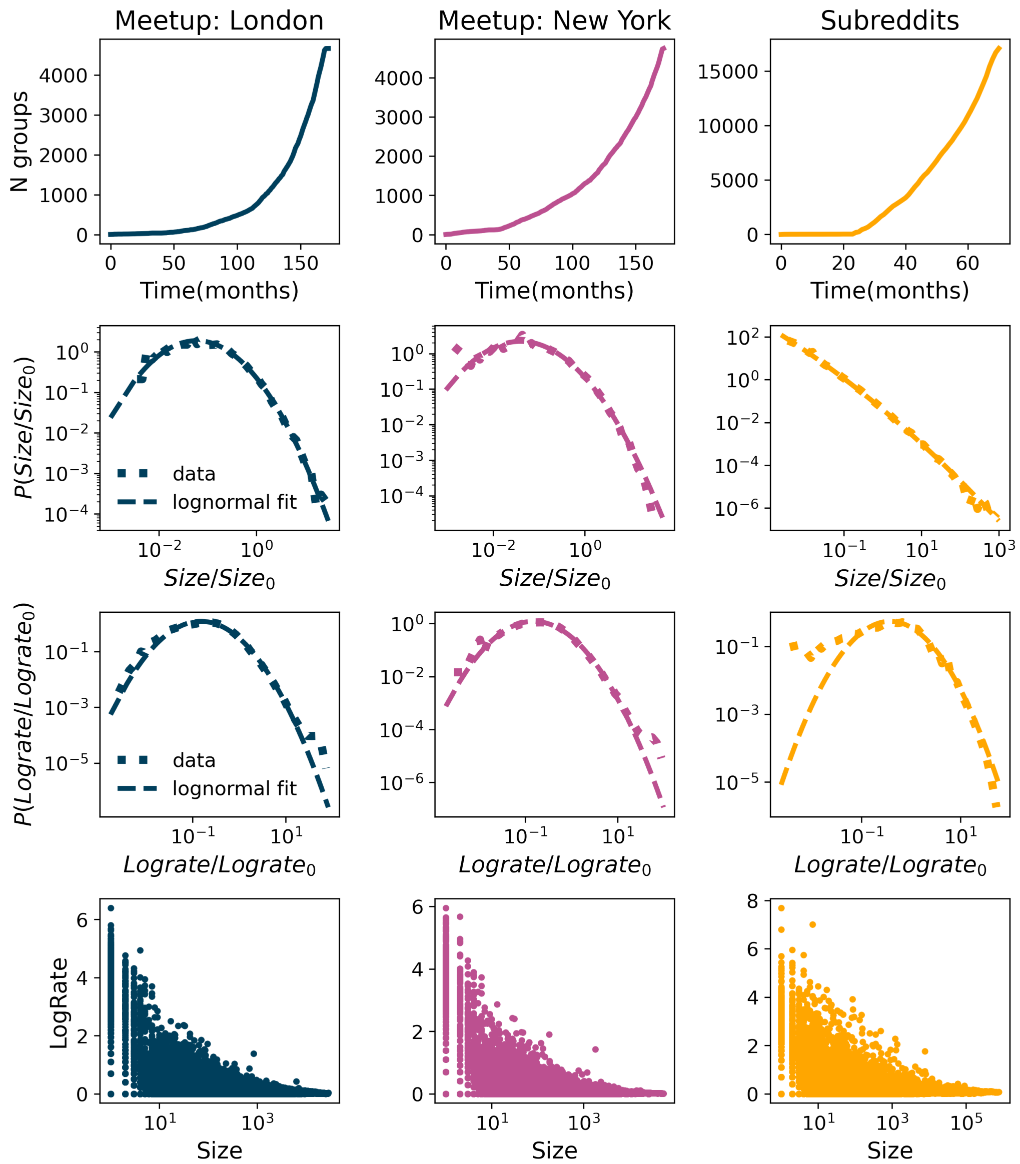

Figure 1.1: The number of groups over time, normalized sizes distribution, normalized log-rates distribution and dependence of log-rates and group sizes for Meetup groups created in London from 08-2002 until 07-2017 that were active in 2017 and subreddits created in the period from 01-2006 to the 12-2011 that were active in 2017.

Even though Meetup and Reddit are different online platforms, we find some common properties of these systems; see figure 1.1. The number of groups and the number of new users grow exponentially. Still, subreddits are larger groups than Meetups. The distribution of groups sizes follows the log-normal distribution:

$$P(S) = \frac{1}{S\sigma\sqrt{2\pi}} exp(-\frac{(\ln(S)-\mu)^2}{2\sigma^2}) \tag{1.1}$$

where S is the group size and μ, and σ are parameters of the distribution. The group sizes distribution of Subreddits is a broad log-normal distribution that resembles the power law. Still, we used the loglikelihood ratio method and showed that log-normal distribution is better than the power-law. More details are given in the result section.

The simplest model that generates the lognormal distribution is multiplicative process [22]. Gibrat used this model to explain the growth of firms. The main assumption of this model is that growth rates $R = log\frac{S_t}{S_{t-\Delta t}}$ do not depend on the size S and that they are uncorrelated. Further, this imply the lognormal distribution of the sizes, while the distribution of growth rates appears to be normal distribution, [31], [32]. Figure 1.2 shows distribution of the lograte, that follow lognormal distribution, contrary to the Gibrat law. Furthermore, lograte depend on the group size 1.2. For these reasons the Gibrat law can not explain the growth of online social groups [33, 34].

The growth of online social groups has universal behavior. It is independent on the size of the group. If we aggregate the groups created in the same year y, and each group size normalize with average size $< S^y >$, $s_i^y = S_i^y / < S^y >$ we will find that group sizes distributions for the same dataset and different years fall on the same line, figure 1.2. The same characteristics are observed for the distribution of the normalized lograte 1.2. The growth is universal in time, and the group sizes distribution do not change from year to year.

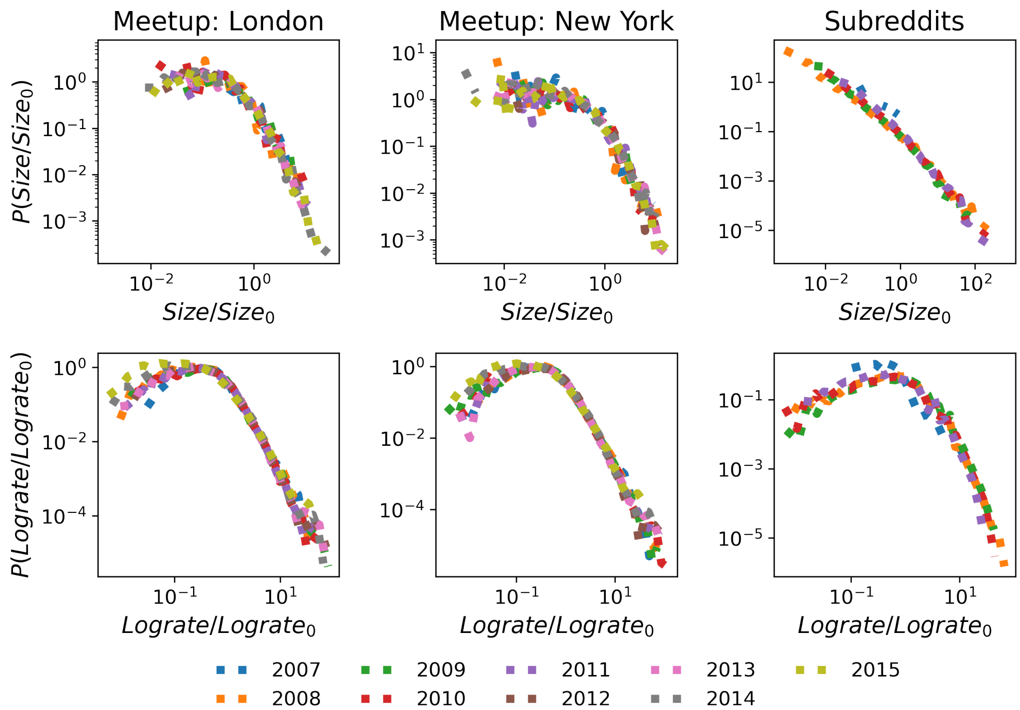

Figure 1.2: The figure shows the groups' sizes distributions and log-rates distributions. Each distribution collects groups founded in the same year and is normalized with its mean value. The group sizes are at the end of 2017 for meetups and 2011 for subreddits.

1.2 The model

Meetup and Reddit engage members in different activities. Still, there are some underlying processes same in both systems. Each member can create new groups and join existing ones. Both systems grow in the number of groups and users, and each user can belong to arbitrary number of groups. In the previous section, we identified the universal patterns in the growth of social groups, but it appears that the growth can not be modelled with Gibrat law.

The complex network models allow us to simulate the growth of these systems considering all types of members' activities. We can identify how model parameters shape the growth process by varying linking rules. Regarding the user's group choice, it was shown that social connections play an important role [35, 36]. On the other hand, users can be driven by personal interests. Diffusion between groups could also be enhanced with rich-get-richer phenomena, where users tend to join larger groups. With a complex network model, we can easily incorporate the nonlinear growth in the number of users and groups, as it is an important parameter that shapes the structure and dynamics of the complex network [37, 38, 39].

The evolution of the social groups has been studied using the co-evolution model in the reference [36]. This model consists of two evolving networks: the bipartite network, which stores connections between users and groups and the affiliation network of social connections. At each time step, active users create new connections in the affiliation network; i.e. they make new friends. They also join existing groups or create new ones, which updates the bipartite network. The group selection can be random with probability proportional to the group size; otherwise, the group is selected through social contacts. Using this model, authors have reproduced the power-law group size distribution found in several communities, such as Flickr or LiveJournal. The empirical analysis of Meetup and Reddit groups showed that group size distribution could be log-normal, meaning that some different mechanisms control the growth of the groups.

We propose a model that is based on the co-evolution model. The main difference between those two models is how model parameters are defined. First of all, in the co-evolution model user becomes inactive after period t_a, which is drawn from an exponential distribution with the rate λ, while in our model probability that the user is active is constant, and the same for each user. The second difference is how groups are chosen. While in the co-evolution model probability that the user selects a group through social linking depends on the friend's degree, we give preference to groups where a user has a larger number of social contacts. We also modified the rules for random linking, so users choose a group with uniform probability.

1.2.1 Groups growth model

The representation of the model is given in figure 1.3. The model consists of two networks:

- bipartite network $\mathcal{B}(V_U, V_G, E_{UG})$, where V_U is set of users, V_G set of groups and E_{UG} set of links between users and groups, where link $e(u, g)$ indicates that user u is member of group g.
- social network $\mathcal{G}(V_U, E_{UU})$ describes the social connections $e(u, v)$ between users u and v, and $V(U)$ is set of users same as in bipartite network.

The bipartite and social networks evolve. At each step, new users $N_U(t)$ are added to the network. It is how the set of users V_U in the bipartite and social network can grow. At arrival, each new member connects to a randomly selected user in the social network G. This allows new members to choose a group based on social contacts [35]. The activity of old members is a stochastic process; old members are activated with probability p_a. The set of active users $\mathcal{A}_U$ has new members $N_U(t)$ and old members who decided to be active in that time step.

The active users can create a new group with probability p_g. By this, group node g is added to the set of group nodes V_G in bipartite network B. If an active user does not create a new group, it will join the existing one with probability $1 - p_g$, see lower panel on figure 1.3. When the user creates a new group or joins an existing one, the link $e(u, g)$ is made in the bipartite network B.

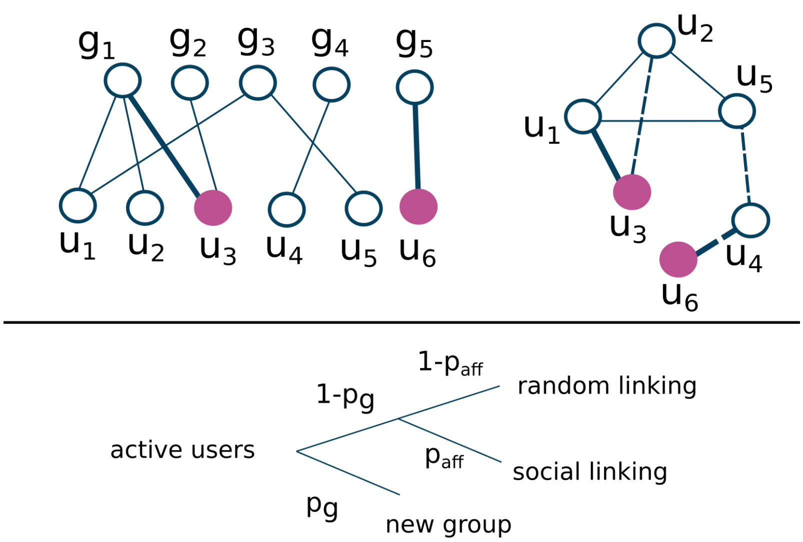

Figure 1.3: The top panel shows bipartite (member-group) and social (member-member) network. Filled nodes are active members, while thick lines are new links in this time step. In the social network dashed lines show that members are friends but still do not share same groups. The lower panel shows model schema, where p_g is probability that user create new group, while p_{aff} is probability that group choice depends on the social connections. **Example:** member u_6 is a new member. First it will make random link with node u_4, and then with probability p_g makes new group g_5. With probability p_a member u_3 is active, while others stay inactive for this time step. Member u_3 will with probability $1 - p_g$ choose to join one of old groups and with probability p_{aff} linking is chosen to be social. As its friend u_2 is member of group g_1, member u_3 will also join group g_1. Joining group g_1, member u_3 will make more social connections, in this case it is member u_1.

When joining existing groups, users may be influenced by social connections. This linking happens with probability p_{aff}. The second case is that the user chooses a random group with probability $1 - p_{aff}$.

Social linking depends on the properties of a bipartite and social network. The networks can be represented with matrices B and A, so if a link between two nodes exists, they have element 1. The neighbourhood of user u, $\mathcal{N}_u$ in a bipartite network is a set of groups in which the user is a member. Similarly, we define the neighbourhood of group g as N_g, as a set of users who belong to the group. From there, we can define the probability P_{ug} that the user u will choose group g. This probability is proportional to the number of social contacts that the user has in the group.

$$P_{ug} = \sum_{u_1 \in \mathcal{N}_g} A_{uu_1} \tag{1.2}$$

After selecting group g, user u is introduced to new members in the group and can make new social contacts. In the simplest case, we could assume that all members belonging to a group are connected. However, previous research on this subject [28, 40, 36] has shown that the existing social connections of members in a social group are only a subset of all possible connections. We select X random members u_i from group g and make new connections in social network $e(u, u_i)$.

The model parameters p_a and p_g are important for controlling the number of users and groups. With larger parameter values p_a, more users become active, and the number of links in bipartite and social networks grows faster. Parameter p_g controls the rate at which new groups are created. For example, if $p_g = 0$, users will not create new groups. Also, if $p_g = 1$, users will only create new groups, and the resulting network will consist of star-like subgraphs. In real systems we do not expect extreme values for probabilities p_a and p_g. First, not all members are constantly active, and we do not find a burst in the creation of the groups. From real data, we notice that there is always a higher number of users than groups in social systems. The parameter p_{aff} how users choose groups, and with higher p_{aff} social connections become more important.

1.2.2 Dependence of the group size distribution on model parameters

Before applying the group growth model on Meetup and Reddit, we consider the system where at each time step, a constant number of users is added $N(t) = 30$. We also fix the probability that user is active to $p_a = 0.1$, so we can in more details explore the influence of parameters p_g and p_{aff}. We plot the group size distribution after 60 steps of simulation. The values of p_g and the p_a influence the number of groups, their maximum size, and the shape of group size distribution. With probability $p_g = 0.1$, users create large number of groups, over 10^4, while with $p_g = 0.5$ they are on the order of magnitude 10^5.

Figure 1.4 show the obtained group size distributions with power-law and log-normal fits. For lower value of parameter $p_g = 0.1$ and $p_{aff} = 0$, users join randomly chosen groups. Group size distributions are approximated with log-normal. When the affiliation parameter is larger, $p_{aff} = 0.5$, the log-normal distribution becomes broader, and so on, we find the larger maximum group size.

If we increase the parameter $p_g = 0.5$, every second active user will create a group. At this group creation rate, the group size distribution deviates from log-normal, but it is not explained with power-law either, right column on figure 1.4.

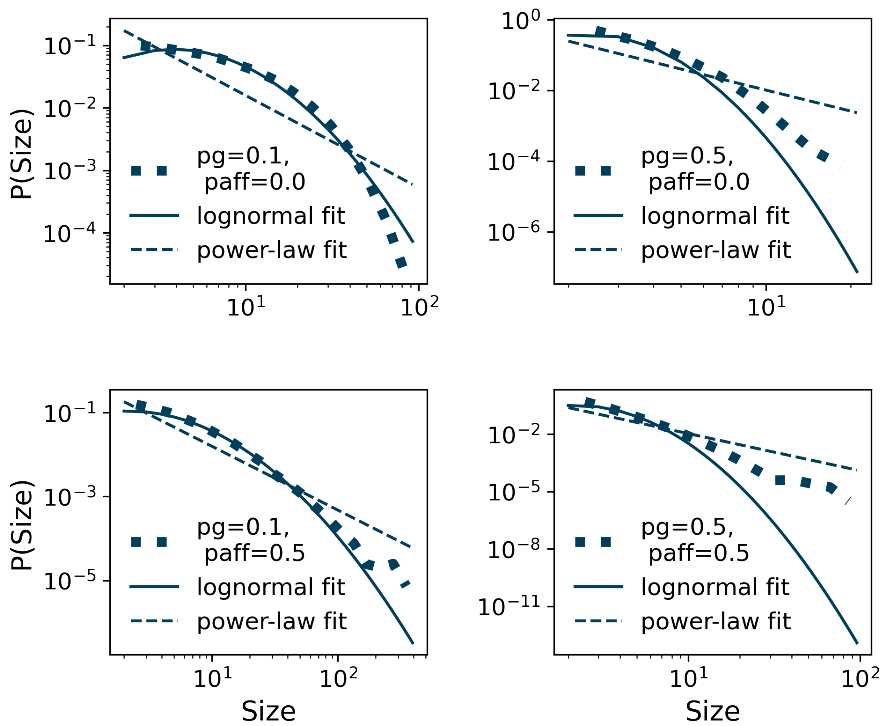

Figure 1.4: The distribution of sizes for different values of p_g and p_{aff} and constant p_a and growth of the system. The combination of the values of parameters of p_g and p_{aff} determine the shape and the width of the distribution of group sizes.

Finally, we compare how group size distribution depends on different rules in random linking. In our model, the probability that the user chooses a random group is uniform. In contrast, in the co-evolution model [36], probability depends on the group size, as in the preferential attachment model.

Instead of random linking, if we incorporate preferential linking, so users with probability $1 - p_{aff}$ tend to choose larger groups, group size distribution changes significantly. Similar to the co-evolution model, we find the power-law distribution. Figure 1.5 shows the results from a model where we add a constant number of new users at each time step. The probabilities p_a and p_g are fixed, and affiliation parameter takes values 0, 0.5 and 0.8. If we consider random linking , top panel on figure 1.5, the distribution becomes broader with larger p_{aff}. On the other hand, with preferential linking, group size distribution is a power-law and the p_{aff} parameter does not have a large impact on the distribution shape.

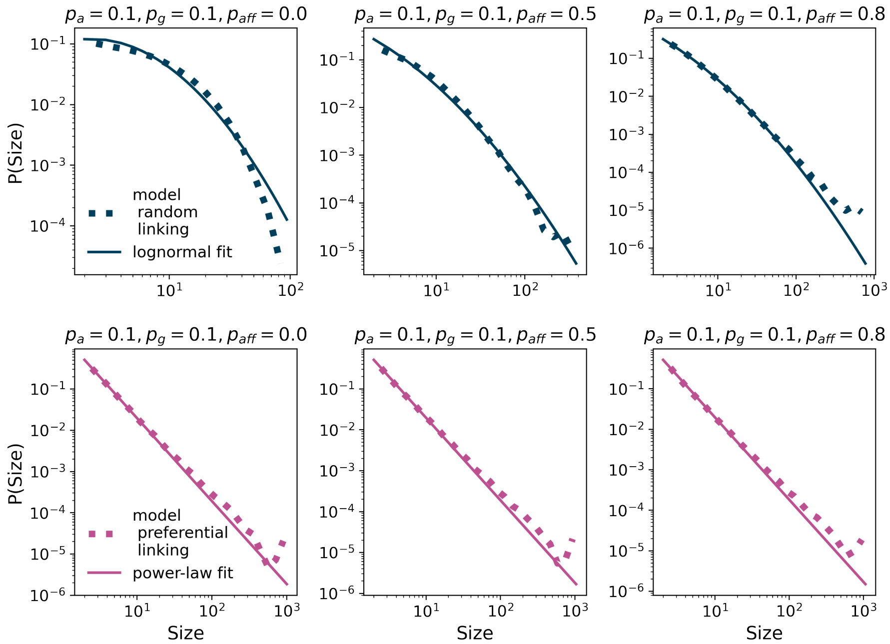

Figure 1.5: Groups sizes distributions for groups model, where at each time step the constant number of users arrive, $N = 30$ and old users are active with probability $p_a = 0.1$. Active users make new groups with probability $p_g = 0.1$, while we vary affiliation parameter p_{aff}. With probability, $1 - p_{aff}$, users choose a group randomly. The group sizes distribution (top row) is described with a log-normal distribution. With higher affiliation parameter, p_{aff}, distribution has larger width. The bottom row presents the case where with probability $1 - p_{aff}$ users have a preference toward larger groups. For all values of parameter p_{aff}, we find the power-law group sizes distribution.

1.3 Results

The social systems do not grow at constant rate. In Ref. [39] authors have shown that features of growth signal influence the structure of social networks. For these reasons we use the real growth signal from Meetup groups located in London and New York, and Reddit community to simulate the growth of the social groups in these systems. Figure 1.6 top panel shows the time series of the number of new members that join each of the three systems each month. All three systems have relatively low growth at the beginning, and than the growth accelerates as the system becomes more popular.

We also use empirical data to estimate p_a, p_g and p_{aff}. Probabilities that old members are active p_a and that new groups are created p_g can be approximated directly from the data. Activity parameter p_a is the ratio between the number of old members that were active in month t and the total number of members in the system at time t. Figure 1.6 middle row shows the variation of parameter p_a during the considered time interval for each system. The values of this parameter fluctuates between 0 and

0.2 for London and New York based Meetup groups, while its value is between 0 and 0.15 for Reddit. To simplify our simulations we assume that p_a is constant in time, and estimate its value as its median value during the 170 months for Meetup systems, and 80 months of Reddit system. For Meetup groups based in London and New York $p_a = 0.05$, while Reddit members are more active on average and $p_a = 0.11$ for this system.

Figure 1.6 bottom row shows the evolution of parameter p_g for the three considered systems. The p_g in month t is estimated as the ratio between the groups created in month t $Ng_{new}(t)$ and the total number of groups that month $Ng_{new}(t) + Ng_{old}(t)$, i.e., $p_g(t) = \frac{Ng_{new}(t)}{N_{new}(t)+N_{old}(t)}$. We see from Fig. 1.6 that $p_g(t)$ has relatively high values at the beginning of the system's existence. This is not surprising. At the beginning these systems have relatively small number of groups and often cannot meet the needs for content of all their members. As the time passes, the number of groups grows, as well as content offerings within the system, and members no longer have a high need to create new groups. Figure 1.6 shows that p_g fluctuates less after the first few months, and thus we again assume that p_g is constant in time and set its value to median value during 170 months for Meetup and 80 months for Reddit. For all three systems p_g has the value of 0.003.

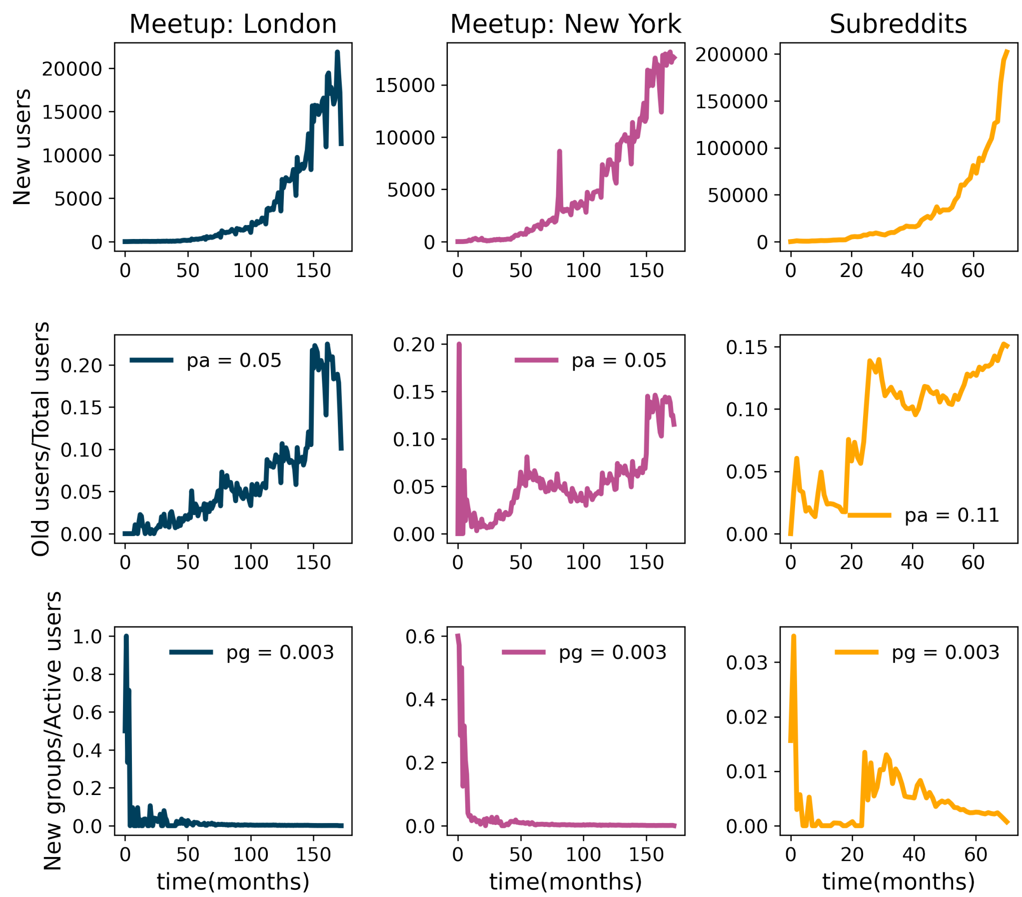

Figure 1.6: The time series of number of new members (top panel), ratio between old members and total members in the system (middle panel), and ratio between new groups and active members(bottom panel) for Meetup groups in London, Meetup groups in New York, and subreddits.

The affiliation parameter p_{aff} is not possible to estimate directly from the empirical data. For these reasons, we simulate the growth of social groups each of the three systems with the time series of new members obtained from the real data and estimated values of parameters p_a and p_g, while we vary the value of p_{aff}. For each of the three systems, we compare the distribution of group sizes obtained from simulations for different values of p_{aff} with ones obtained from empirical analysis using Jensen Shannon (JS) divergence. The JS divergence [41] between two distributions P and Q is defined as

$$JS(P,Q) = H\left(\frac{P+Q}{2}\right) - \frac{1}{2}(H(P) + H(Q)) \tag{1.3}$$

where $H(p)$ is Shannon entropy $H(p) = \sum_x p(x)log(p(x)$. The JS divergence is symmetric and if P is identical to Q, $JS = 0$. The smaller the value of JS divergence, the better is the match between empirical and simulated group size distributions. The Table 1.1 shows the value of JS divergence for all three systems. We see that for London based Meetup groups the affiliation parameter is $p_{aff} = 0.5$, for New York groups $p_{aff} = 0.4$, while the affiliation parameter for Reddit $p_{aff} = 0.8$. Our results show that social diffusion is important in all three systems. However, Meetup members are more likely to join groups at random, while for the Reddit members their social connections are more important when it comes to choice of the subreddit.

p_{aff}	JS cityLondon	JS cityNY	JS reddit2012
0.1	0.0161	0.0097	0.00241
0.2	0.0101	0.0053	0.00205
0.3	0.0055	0.0026	0.00159
0.4	0.0027	**0.0013**	0.00104
0.5	**0.0016**	0.0015	0.00074
0.6	0.0031	0.0035	0.00048
0.7	0.0085	0.0081	0.00039
0.8	0.0214	0.0167	**0.00034**
0.9	0.0499	0.0331	0.00047

Table 1.1: Jensen Shannon divergence between group sizes distributions from model (in model we vary affiliation parameter paff) and data.

Figure 1.7 shows the comparison between the empirical and simulation distribution of group sizes for three considered systems. We see that empirical distributions for Meetup groups based in London and New York are perfectly reproduced by the model and chosen values of parameters. In the case of Reddit, the distribution is very broad, and the tail of distribution is well reproduced by the model. The bottom row of Fig. 1.7 shows the distribution of logarithmic values of growth rates of groups obtained from empirical and simulated data. We see that the tails of empirical distributions for all three systems are well emulated by the ones obtained from the model. However, there are deviations which are the most likely consequence of using median values of parameters p_a, p_g, and p_{aff}.

1.4 Distributions fit

We compute the log-likelihood ratio R, and p-value between different distributions and log-normal fit [42] to determine the best fit for the group size distributions. Distribution with a higher likelihood is a better fit. The log-likelihood ratio R then has a positive or negative value, indicating which distribution represents a better fit. To choose between two distributions, we need to calculate p-value, to be sure that R is sufficiently positive or negative and that it is not the result of chance fluctuation from the result that is close to zero. If the p-value is small, $p < 0.1$, it is unlikely that the sign of R is the chance of fluctuations, and it is an accurate indicator of which model fits better.

Table 1.2 summarizes the findings for empirical data on group size distributions from Meetup groups in London, Meetup groups in New York and Reddit. Using the maximum likelihood method,

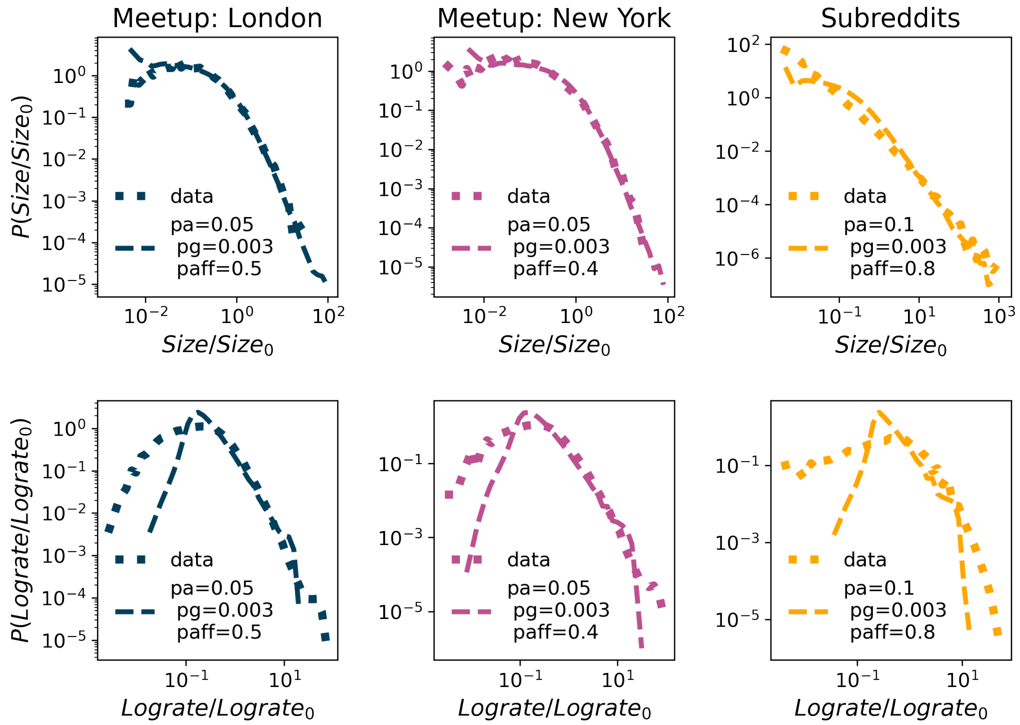

Figure 1.7: The comparison between empirical and simulation distribution for group sizes (top panel) and lograte s (bottom panel).

we obtain the parameters of the distributions [43]. The results indicate that log-normal distribution is the best fit for all three systems. Figure 1.8 shows the distributions of empirical data as well as log-normal fit on data. For Meetup data, we present fit on stretched exponential distribution, which very well fits a large portion of data. For subreddits, distribution is broad and, potentially, resembles power-law. Still, log-normal distribution is a more suitable fit.

Table 1.2: The likelihood ratio R and p-value between different candidates and **lognormal** distribution for fitting the distribution of **groups sizes** of Meetup groups in London, New York and in Reddit. According to these statistics, the lognormal distribution represents the best fit for all communities.

distribution	Meetup city London		Meetup city NY		Reddit	
	R	p	R	p	R	p
exponential	-8.64e2	8.11e-32	-8.22e2	6.63e-26	-3.85e4	1.54e-100
stretched exponential	-3.01e2	1.00e-30	-1.47e2	7.78e-8	-7.97e1	5.94e-30
power law	-4.88e3	0.00	-4.57e3	0.00	-9.39e2	4.48e-149
truncated power law	-2.39e3	0.00	-2.09e3	0.00	-5.51e2	2.42e-56

We use the same methods to estimate the fit for simulated group size distributions on Meetup groups in London, New York, and Subreddits. Table 1.3 shows the results of the log-likelihood ratio R and p-value between different distributions. We conclude that log-normal distribution is most suitable for simulated group size distributions. Plotting log-normal and stretched exponential fit on data, Fig. 1.9 we confirm our observations.

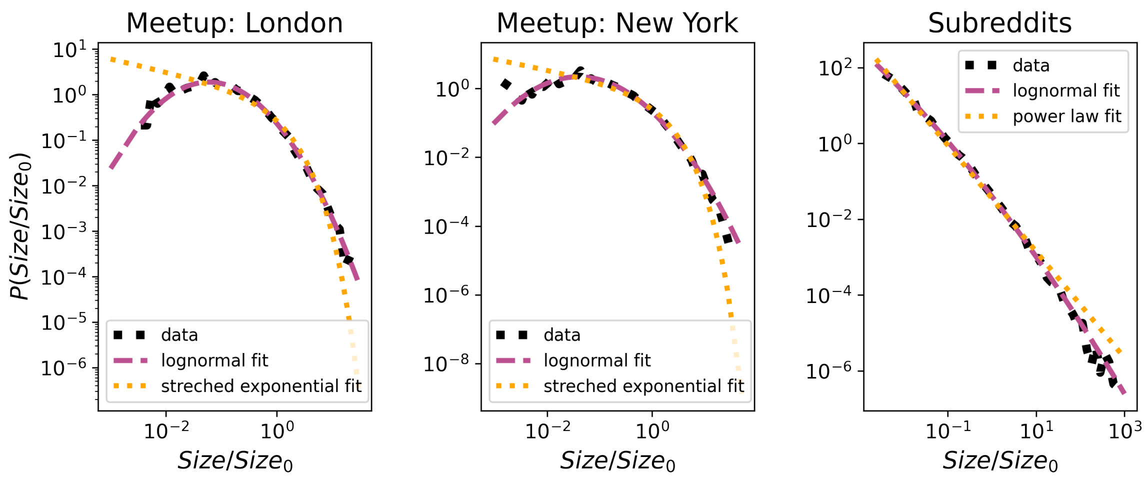

Figure 1.8: The comparison between log-normal and stretched exponential fit to London and NY data, and between log-normal and power law for Subreddits. The parameters for log-normal fits are 1) for city London $\mu = -0.93$ and $\sigma = 1.38$, 2) for city NY $\mu = -0.99$ and $\sigma = 1.49$, 3) for Subreddits $\mu = -5.41$ and $\sigma = 3.07$.

Table 1.3: The likelihood ratio R and p-value between different candidates and **lognormal** distribution for fitting the distribution of **simulated group sizes** of Meetup groups in London, New York and Reddit. According to these statistics, the lognormal distribution represents the best fit for all communities.

distribution	Meetup city London		Meetup city NY		Reddit	
	R	p	R	p	R	p
exponential	-6.27e4	0.00	-5.11e4	0.00	-1.26e5	7.31e-125
stretched exponential	-1.01e4	1.96e-287	-6.69e3	1.46e-93	-1.39e4	0.00
power law	-2.29e5	0.00	-3.73e5	0.00	-4.38e4	0.00
truncated power law	-9.28e4	0.00	-1.55e5	0.00	-9.12e4	0.00

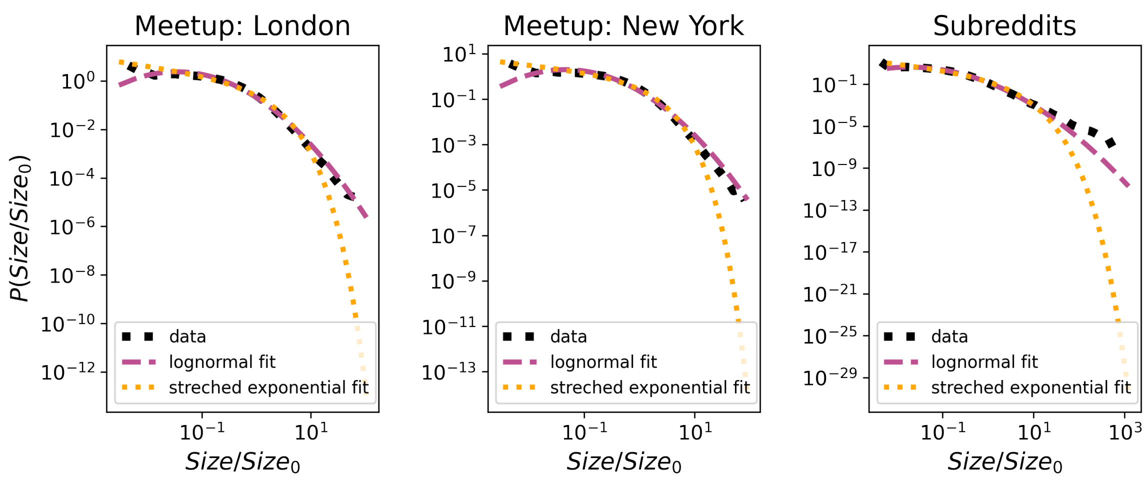

Figure 1.9: The comparison between lognormal and stretched exponential fit to simulated group sizes distributions. The parameters for log-normal fits are 1) for city London $\mu = -0.97$ and $\sigma = 1.43$, 2) for city NY $\mu = -0.84$ and $\sigma = 1.38$, 3) for Subreddits $\mu = -1.63$ and $\sigma = 1.53$.

Bibliography

[1] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

[2] V. Latora, V. Nicosia, and G. Russo. Complex networks: Principles, methods and applications. 2017.

[3] L da F Costa, Francisco A Rodrigues, Gonzalo Travieso, and Paulino Ribeiro Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in physics*, 56(1):167–242, 2007.

[4] Luciano da Fontoura Costa, Osvaldo N Oliveira Jr, Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana, and Luis Enrique Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.

[5] Guido Caldarelli. *Scale-Free Networks: Complex Webs in Nature and Technology*. 05 2007.

[6] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016.

[7] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[8] Petter Holme and Jari äki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.

[9] Naoki Masuda and Renaud Lambiotte. *A Guide to Temporal Networks*. 10 2016.

[10] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):1–30, 2015.

[11] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1):1–16, 2012.

[12] Naomi A Arnold, Benjamin Steer, Imane Hafnaoui, Hugo A Parada G, Raul J Mondragon, Félix Cuadrado, and Richard G Clegg. Moving with the times: Investigating the alt-right network gab with temporal interaction graphs. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–17, 2021.

[13] Mason A Porter. What is... a multilayer network. *Notices of the AMS*, 65(11), 2018.

[14] Alberto Aleta and Yamir Moreno. Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics*, 10(1):45–62, 2019.

[15] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 07 2014.

[16] Tiago A Schieber, Laura Carpi, Albert Díaz-Guilera, Panos M Pardalos, Cristina Masoller, and Martín G Ravetti. Quantification of network structural dissimilarities. *Nature communications*, 8(1):1–10, 2017.

[17] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1 – 44, 2016. Community detection in networks: A user guide.

[18] Martin Rosvall, Jean-Charles Delvenne, Michael T. Schaub, and Renaud Lambiotte. Different approaches to community detection. *CoRR*, abs/1712.06468, 2017.

[19] Ryan J Gallagher, Jean-Gabriel Young, and Brooke Foucault Welles. A clarified typology of core-periphery structure in networks. *arXiv preprint arXiv:2005.10191*, 2020.

[20] Kamalika Basu Hajra and Parongama Sen. Phase transitions in an aging network. *Physical Review E*, 70(5):056103, 2004.

[21] Thorben Funke and Till Becker. Stochastic block models: A comparison of variants and inference methods. *PLOS ONE*, 14(4):1–40, 04 2019.

[22] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet mathematics*, 1(2):226–251, 2004.

[23] J. Nair, A. Wierman, and B. Zwart. *The Fundamentals of Heavy Tails*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2022.

[24] Jan W Kantelhardt, Stephan A Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.

[25] Espen Alexander Fürst EAFI Ihlen. Introduction to multifractal detrended fluctuation analysis in matlab. *Frontiers in physiology*, 3:141, 2012.

[26] A. Melnikov, J. Lee, V. Rivera, M. Mazzara, and L. Longo. Towards dynamic interaction-based reputation models. In *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, pages 422–428, 2018.

[27] Sergey N Dorogovtsev and José FF Mendes. Scaling properties of scale-free evolving networks: Continuous approach. *Physical Review E*, 63(5):056125, 2001.

[28] Jelena Smiljanić and Marija Mitrović Dankulov. Associative nature of event participation dynamics: A network theory approach. *PloS one*, 12(2):e0171565, 2017.

[29] Milovan Suvakov, Marija Mitrovic, Vladimir Gligorijevic, and Bosiljka Tadic. How the online social networks are used: dialogues-based structure of myspace. *Journal of The Royal Society Interface*, 10(79):20120819, 2013.

[30] Hernán A Makse, Shlomo Havlin, Moshe Schwartz, and H Eugene Stanley. Method for generating long-range correlations for large systems. *Physical Review E*, 53(5):5445, 1996.

[31] Hernan Mondani, Petter Holme, and Fredrik Liljeros. Fat-tailed fluctuations in the size of organizations: the role of social influence. *PLoS One*, 9(7):e100527, 2014.

[32] Dongfeng Fu, Fabio Pammolli, Sergey V Buldyrev, Massimo Riccaboni, Kaushik Matia, Kazuko Yamasaki, and H Eugene Stanley. The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences*, 102(52):18801–18806, 2005.

[33] Gerald F Frasco, Jie Sun, Hernán D Rozenfeld, and Daniel Ben-Avraham. Spatially distributed social complex networks. *Physical Review X*, 4(1):011008, 2014.

[34] Jiang-Hai Qian, Qu Chen, Ding-Ding Han, Yu-Gang Ma, and Wen-Qing Shen. Origin of gibrat law in internet: Asymmetric distribution of the correlation. *Physical Review E*, 89(6):062808, 2014.

[35] Sanjay Ram Kairam, Dan J Wang, and Jure Leskovec. The life and death of online groups: Predicting group growth and longevity. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 673–682, 2012.

[36] Elena Zheleva, Hossam Sharara, and Lise Getoor. Co-evolution of social and affiliation networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1007–1016, 2009.

[37] Marija Mitrović, Georgios Paltoglou, and Bosiljka Tadić. Quantitative analysis of bloggers' collective behavior powered by emotions. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(02):P02005, 2011.

[38] Marija Mitrović Dankulov, Roderick Melnik, and Bosiljka Tadić. The dynamics of meaningful social interactions and the emergence of collective knowledge. *Scientific reports*, 5(1):1–10, 2015.

[39] Ana Vranić and Marija Mitrović Dankulov. Growth signals determine the topology of evolving networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(1):013405, 2021.

[40] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, 2006.

[41] Jop Briët and Peter Harremoës. Properties of classical and quantum jensen-shannon divergence. *Phys. Rev. A*, 79:052311, May 2009.

[42] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[43] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: A python package for analysis of heavy-tailed distributions. *PLOS ONE*, 9(1):1–11, 01 2014.

[44] Akrati Saxena and Harita Reddy. Users roles identification on online crowdsourced q&a platforms and encyclopedias: a survey. *Journal of Computational Social Science*, pages 1–33, 2021.

[45] Tiago Santos, Simon Walk, Roman Kern, Markus Strohmaier, and Denis Helic. Activity archetypes in question-and-answer (q8a) websites—a study of 50 stack exchange instances. *ACM Transactions on Social Computing*, 2(1):1–23, 2019.

[46] Rogier Slag, Mike de Waard, and Alberto Bacchelli. One-day flies on stackoverflow-why the vast majority of stackoverflow users only posts once. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*, pages 458–461. IEEE, 2015.

[47] Anamika Chhabra and S RS Iyengar. Activity-selection behavior of users in stackexchange websites. In *Companion Proceedings of the Web Conference 2020*, pages 105–106, 2020.

[48] Himel Dev, Chase Geigle, Qingtao Hu, Jiahui Zheng, and Hari Sundaram. The size conundrum: Why online knowledge markets can fail at scale. In *Proceedings of the 2018 World Wide Web Conference*, pages 65–75, 2018.

[49] Tiago Santos, Simon Walk, Roman Kern, Markus Strohmaier, and Denis Helic. Self-and cross-excitation in stack exchange question & answer communities. In *The World Wide Web Conference*, pages 1634–1645, 2019.

[50] Chiara Orsini, Marija M Dankulov, Pol Colomer-de Simón, Almerima Jamakovic, Priya Mahadevan, Amin Vahdat, Kevin E Bassler, Zoltán Toroczkai, Marián Boguná, Guido Caldarelli, et al. Quantifying randomness in real networks. *Nature communications*, 6(1):1–10, 2015.

[51] Damon Centola, Víctor M Eguíluz, and Michael W Macy. Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications*, 374(1):449–456, 2007.

[52] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[53] Ekaterina Yashkina, Arseny Pinigin, JooYoung Lee, Manuel Mazzara, Akinlolu Solomon Adekotujo, Adam Zubair, and Luca Longo. Expressing trust with temporal frequency of user interaction in online communities. Advances in Intelligent Systems and Computing, pages 1133–1146, Cham, 2020. Springer International Publishing.

[54] Yaniv Dover, Jacob Goldenberg, and Daniel Shapira. Sustainable online communities exhibit distinct hierarchical structures across scales of size. *Proceedings of the Royal Society A*, 476(2239):20190730, 2020.

[55] Vincent Labatut and Hocine Cherifi. Accuracy measures for the comparison of classifiers. *arXiv preprint arXiv:1207.3790*, 2012.

[56] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*, 2005(09):P09008, 2005.

[57] Jorge M Santos and Mark Embrechts. On the use of the adjusted rand index as a metric for evaluating supervised classification. In *International conference on artificial neural networks*, pages 175–184. Springer, 2009.

[58] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.