



From Generality to Specificity: On Matter of Scale in Social Media Topic Communities

Danila Vaganov^(✉), Mariia Bardina, and Valentina Guleva^(✉)

ITMO University, 49 Kronverksky Pr., St. Petersburg 197101, Russia
{vaganov,guleva}@itmo.ru

Abstract. Research question stated in current paper concerns measuring significance of interest topic to a person on the base of digital footprints, observed in on-line social media. Interests are represented by on-line social groups in VK social network, which were marked by topics. Topic significance to a person is supposed to be related to the fraction of representative groups in user's subscription list. We imply that for each topic, depending on its popularity, relation to geographical region, and social acceptability, there is a value of group size which is significant. In addition, we suppose, that professional clusters of groups demonstrate relatively higher inner density and unify common groups. Therefore, following groups from more specific clusters indicate higher personal involvement to a topic – in this way, representative topical groups are marked. We build social group similarity graph, which is based on the number of common followers, extract subgraphs related to a single topic, and analyse bins of groups, build with increase of group sizes. Results show topics of general interests have higher density at larger groups in contrast to specific interests, which is in correspondence with initial hypothesis.

Keywords: Topic communities · On-line social media · Similarity group network · Personal involvement · General interest · Specific interest · Scaling phenomena

1 Introduction

Interests play a big role in people's lives [22]. They are influenced by our life as well as they are causing influence on it. In psychological studies, researchers usually explore development of interest in relation to academic field and career path. They examine how interest can affect motivation in order to understand how to increase engagement in studying. Rotgans and Schmidt [26] propose a model of interest formation that focuses on how prior relationship to the topic

This research is financially supported by The Russian Science Foundation, Agreement #17-71-30029 with co-financing of Bank Saint Petersburg.

© Springer Nature Switzerland AG 2020
V. V. Krzhizhanovskaya et al. (Eds.): ICCS 2020, LNCS 12140, pp. 305–318, 2020.
https://doi.org/10.1007/978-3-030-50423-6_23

affects formation of interest and discuss how interest can emerge when person is confronted with a problem. Harackiewicz and Knogler emphasize how future choices and career path can be influenced by interests in all stages of development [12]. The advertisement industry is also interested in finding what is behind an interest. They aim at determining current personal interests and patterns of their possible change. Modern day involvement of people in social media is high, almost 3.5 billion people actively use social media around the world and over 70 million people in Russia¹, which makes data, collected from social media, a great asset in Social studies.

Personal interest analysis, performed by means of social media data, requires estimation of group importance as a marker of topic interest and involvement. Big groups of general interest attract people of different preferences, education, and occupation. Groups presenting specific context are less related to our daily life. In this way, they can imply entry threshold, restricting maximal number of people involved [2]. Therefore, topic popularity affects the upper threshold of group size.

In addition to the effects of participation costs, there exist an influence of topic consistency cues. This is promoted by attraction-selection-attrition (ASA) theory [2], which concerns factors, attracting people to communities, in the model of community evolution. In this way, they concern both, group size related to its age and ability of people to get content and be satisfied.

Other aspect, characterising personal involvement is the number of groups related to a topic and containing similar sets of members. Let think about critical case of such a phenomena as a professional community. In this way, we can formalise “specificity” of groups union if they are intersected enough and related to similar topics. Then, relation to a topic with strong intersections and number of groups from the cluster is implied to be an indicator of personal involvement.

This assumption is reinforced by Cinelli et al. [5], showing experts being active in several groups, while majority of people are satisfied by following outstanding one.

In this study we are aimed at exploration the groups of different topics and sizes to obtain patterns, being able to characterise personal involvement to a topic on the base of subscriptions to social groups. We suppose that there are values of group sizes, which are significant to a certain topic. For this purpose we consider group similarity graph with edges, weighted as group intersection, divide groups between bins of different sizes, and explore density of a topic cluster relatively to the rest of the topics. Similarly, we fix a topic and explore densities related to interscale densities. This allows for distinguishing the most connected clusters and corresponding group size, and mitigate effects of topic popularity. In this way, we conclude that the smallest groups related to the most dense topic clusters indicate higher personal involvement to the topic.

¹ <https://datareportal.com/reports/digital-2019-russian-federation>.

2 Related Literature

Categorisation of interests is mainly related to personal energy associated with a topic, an active personal concern for a certain object or activity [7]. Interests can be triggered by an event (situational) or remain stable over considered amount of time (individual) [25], like emotions and feelings. Interests development from situational to individual can be considered as four phase process [13, 24], depending on interest “stability”.

The evolution of interests is driven by perceived values [14], which is related to subject utility, and can be reasoned by both, social [18] and personal values. Possibility of communication with other people triggers interest development [30], as well as sense of social belonging, showing increase in interest and involvement of people working in groups in contrast to people working alone [3]. In addition, acceptability of interest may result in interest formation [15].

Interest acceptability is strongly related to personal background. Social status, gender, age, and other features, affecting personal perception, sensibility, and ability to be focused form initial base and background for further scope of interests. Women are less engaged in science, engineering, and math [4], pupil of different age and socioeconomic status were shown to differ in music interests [19], age and gender influence interest in learning about cooking [35], factors like school location (urban or rural) and racial composition affect differences in gender gap and arts consumption patterns [27].

Applied studies are related to the exploration of existing structures in friendship networks, e.g. user segmentation approaches [6], and their relation to personal interests [16]. In this way, correlation between interest similarity and demographic factors (gender, age, and location) is studied [11]. To measure similarity of interests, the relative number of common interests can be taken. Authors take how rare common interests are among the general population [11]. Guleva et al. concluded, that topology of friendships in a social group significantly depends on the topic, particularly in terms of degree assortativity and clustering coefficient [10]. Backstrom et al. [1] show that the transitivity of friendship (clustering coefficient), being in common community, affect the decision to join them; not only absolute number of friends there. Correlation between structure of interactions and interests topic category [33] show, that topics can be ordered according to their social importance.

Other class of applied studies are related to categories of interest prediction [36], usage of recommending systems [9, 17], and topic hierarchy trees [17], like Wikipedia graphs [8, 23]. Wikipedia graphs are also used to build user interest profiles. Modelling of short- and long-term interests uses neural networks on the base of click-streams [29].

3 Data Set and Preprocessing

We gather profiles of 28,520 groups from social media [VK.com](https://vk.com), selecting groups which subscribers are self-affiliated with Saint-Petersburg. For each group we

collected captions, brief description (up to 200 words), status, list of followers and up to 20 textual posts. During the preprocessing, we combined group captions, statuses, descriptions, and collected posts into a single document. Then we remove all special symbols, numbers, non-cyrillic words, and stop-words.

For each group, we assigned the number of followers by measuring length of each collected follower list. To perform separation into the levels, we used the logarithmic binning and divided the whole data-set of groups into $p = 10$ levels (number of bins), where, for bin of order i , the size s_i is defined as follows:

$$s_i = \min(L) \cdot \frac{\max(L)^{\frac{i-1}{p}}}{\min(L)}, \tag{1}$$

where L is the set of sizes of all considered groups in the data set. It is important to emphasize, that further in the paper we refer to a certain bin of group size as the term “level”.

Figure 1 reflects the distribution of group sizes and corresponding binning, i.e. separation on levels. One can observe that this fits power-law distribution, and in this work we consider the groups with sizes from several hundreds to tens of millions subscribers. The highest fraction of groups is placed on the 5th level, where group size varies from 54 K to 160 K. Levels 3 and 6 are also prevalent.

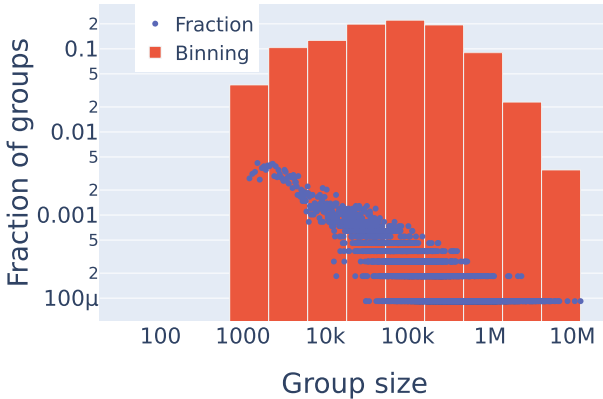


Fig. 1. Distribution of the group size and obtained log bins (levels)

4 Method

In this study, we consider a case of scaling phenomena observed in on-line social groups, also called subscriptions, which represent internet communities related to a certain topic of interest. In this way, for each user, a group list reflects their involvement into different topics. We assume that the scale, corresponding to

group size reflects a lot of characteristics such as group popularity or specificity of particular interest. To explore these scales separately we apply logarithm binning and extract several levels of groups. For the extracted levels we build similarity network, demonstrating their collaborative relation, i.e. how strong is the tendency of users to prefer each pair of groups, and do topic modeling to build a connection with characteristics of extracted levels.

4.1 Building Similarity Network of Groups

Similarity between two groups is taken as normalised intersection between their subscribers. To compute similarity θ between groups $\gamma_i, \gamma_j \in \Gamma$, we consider the cosine measure of the corresponding subscriber sets $V(\gamma_i), V(\gamma_j)$:

$$\theta(\gamma_i, \gamma_j) = \frac{|V(\gamma_i) \cap V(\gamma_j)|}{\sqrt{|V(\gamma_i)| \cdot |V(\gamma_j)|}} \quad (2)$$

After the calculation of similarity between all groups, one obtain a weighted complete graph, where nodes correspond to groups and edge weights reflect similarity between them. In order to separate it into the clusters of closest groups, one should find the edge between degree of separation and connectivity of clusters as described in [32]. We vary a threshold of the lowest possible similarity between groups and look for the best intercluster separation.

Then, for each group we describe a prevailing topic, using topic modelling techniques, aimed at comparison of similarity networks and topics at different scales. The final goal is to build interconnection between scales and to describe semantic differences.

4.2 Topic Modelling

Topic modelling is performed to extract key words, describing group topic. Posts published by a group are collected in a document. After that, lemmatization on the set of collected documents is performed by means of a morphological analyzer Mystem [28]. For topic modeling we use Additive Regularization of Topic Models (ARTM) [34], model implemented in BigARTM library². Key feature of this model is the ability to assign combination of regularizers (a criterion to be maximized) for better model tuning. To train a model we use combination of two regularizers, `SmoothSparseThetaRegularizer` and `DecorrelatorPhiRegularizer`. First regularizer is responsible for smoothing or sparsing topics. The second one provides decorrelation of topics, which is needed to make the learned topics more interpretable. Both regularizers are controlled by the coefficients of regularization (τ_1 and τ_2 , respectively). The optimal number of topics and values of regularizers are chosen based on perplexity and coherence measure. The perplexity measure indicates the level and speed of convergence of the model. It is defined as

$$\mathcal{P} = \exp \left(-\frac{1}{n} \sum_{d \in D} \sum_{w \in d} n_{dw} \ln p(w|d) \right), \quad (3)$$

² BigARTM open source project: bigartm.org.

where D is set of documents, n_{dw} is frequency of word w in document d , $p(w|d)$ is the probability of a term w to occur in a document d . The coherence measure is well correlated with human evaluation of interpretability [20] and is defined as

$$C_t = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k value(w_i, w_j), \tag{4}$$

where k is a number of the most representative tokens for the topic and $value$ is pairwise information about tokens, for example, as used in [20], the pointwise mutual information (PMI) is:

$$value(w_i, w_j) = PMI(w_i, w_j) = \left[\log \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \right], \tag{5}$$

which is used to measure the similarity of words w_i and w_j based on co-occurrence statistics.

4.3 Topic Density on a Certain Level

The main idea of our study is to measure how strong group connections are inside a certain topic in relation to the groups outside the topic on a certain level of group sizes, an example illustrated on (Fig. 2).

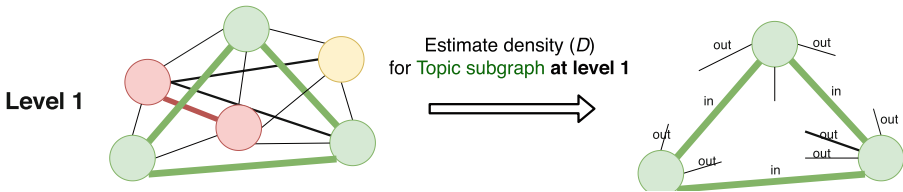


Fig. 2. Illustration of a subgraph extraction for evaluation of topic relative density at level 1.

Formally the relative density for a given subgraph is described by equation:

$$D = \frac{\sum_{e \in E_T} w(e) / (N_T \cdot (N_T - 1))}{\sum_{e \in E \setminus E_T \setminus E_{\bar{T}}} w(e) / (N_T \cdot N_{\bar{T}})}, \tag{6}$$

where $w(e)$ is weight of an edge e , showing similarity between groups; N_T is the number of nodes related to topic T ; numerator is weighted network density for groups, related to topic T , and denominator reflects weighted sum of edges between T and all other group topics, related to all possible edge weights between them. Maximal edge weight is supposed to be 1 (due to Eq. (2)). For this measure we consider each level separately.

5 Results

5.1 Similarity Graph

Following Sect. 4.1, we use groups, having at least 10 posts, to construct a similarity graph. A calculated optimal weight threshold is $\theta \leq 0.066$, which guarantees that all groups are in the same connected component. As a result, the number of nodes is 12,092 and number of edges is 917K.

5.2 Topic Modelling

After lemmatization, we removed all documents consisting of less than 5 lemmas. To determine the optimal number of topics, we trained ARTM model without regularizers with number of passes set to 20. As a result, the number of topics varied from 20 to 110 with an increment of 5 as presented in Fig. 3. Based on coherence measure, the optimal number of topics was set to 80. For training with regularizers, which were described in Sect. 4.2, we conducted an experiment where τ_1 varied from -0.05 to -0.4 with step 0.05 and τ_2 varies from $2 \cdot 10^4$ to $16 \cdot 10^4$ with step $2 \cdot 10^4$. The best result, obtained during training with different variables of regularizers, was achieved at $\tau_1 = -0.2$ and $\tau_2 = 10^5$.

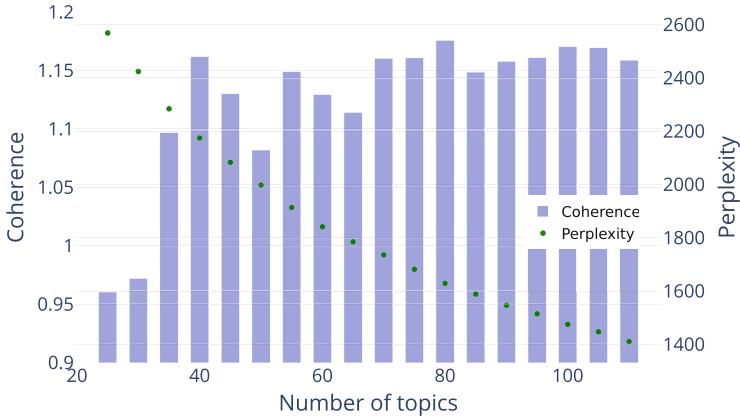


Fig. 3. Results of the experiment for finding right number of topics

After that, we gathered 80 topics represented by top-5 most probable words in each topic. On the base of that representation, each topic was manually assigned by keywords. Then, the topics marked as “noise” were removed. After that, we assigned the most probable topics to each group. To leave only representative groups, we need to choose a threshold for the value of probability, according to which topics were assigned to the groups. Each group, having the probability of being attributed to the topic less than the value of threshold, was excluded.

We looked into the distribution of the number of groups in a topic and set a value of threshold to 0.4. The chosen distribution and its relationship to the original set of groups is shown in Fig. 4.

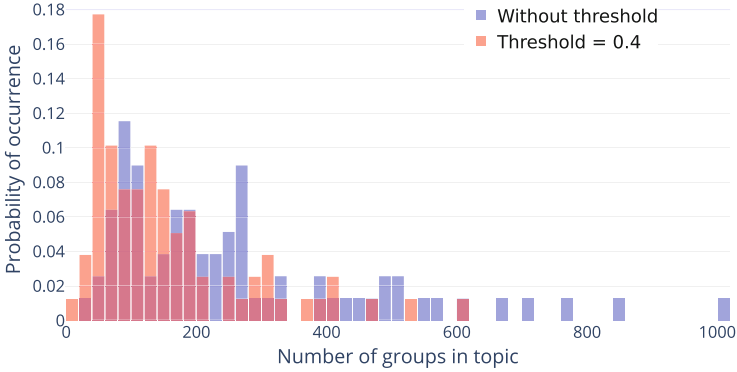


Fig. 4. Distribution of the number of groups in topic with and without threshold set up

After filtering of (non-representative) topics, showing less than 20 related groups, we obtain 28 topics, containing more than 7K groups. In the rest of the study we consider similarity subgraph for the chosen groups.

5.3 Relation Between Topics and Group Preference

In this section we begin with the exploration of the interrelation between the obtained topics and the graph of groups similarity, and then we investigate an influence of group sizes.

First, consider an illustration of giant component of obtained graph, where the color of node corresponds to group topic (Fig. 5). For better representation, we prune weak ties by the weight, with threshold of similarity $\theta \leq 0.1$, and remove some groups, which are not in giant component. One can observe, that groups of the same topics form natural clusters based on the similarity of users’ subscriptions. Some topics form independent weakly connected modules, for instance, a blue cluster in the right part of the figure corresponds to the “furniture” topic, while huge pink module is “handcraft” topic.

To estimate a strength of relation between topics and overlapping group affiliation numerically, we calculate, for each topic $t_i \in T$, the maximum intersection between groups from each topic ($g_t \in t_i$) and groups in graph clusters ($g_c \in c_j$) as:

$$I_{max_i} = \max\left(\frac{|t_i \cap c_j|}{|t_i|}\right), c_j \in C \tag{7}$$

where clusters ($c_j \in C$) were obtained by means of the Leiden algorithm [31] and an optimal value of modularity (defined by Newman et al. [21]) is 0.54,

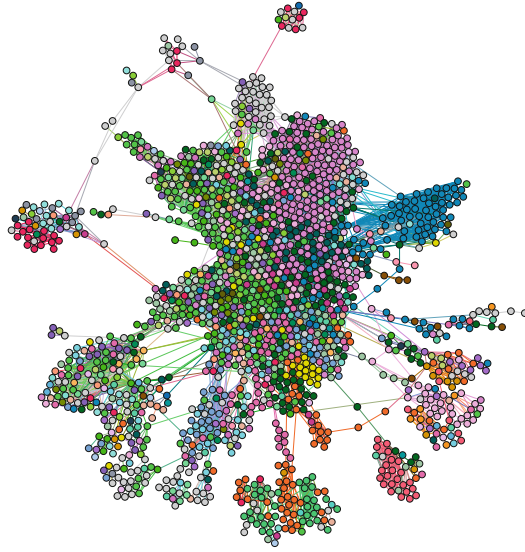


Fig. 5. The obtained graph of groups similarity (nodes colored by topics).

which means that considered graph is well separated (i.e. groups forms a certain blocks by interests). As a result, we obtain 38 clusters. In Fig. 6 one can observe the values of maximum intersections I_{max} obtained for some topics. As we have seen before, the topic cluster “furniture” gets the highest intersection with one of the modularity clusters and close to the maximum. In contrast, “activity in Saint-Petersburg” reaches the minimal value of intersection: the possible reason is that different activities can be associated with different topics, but not in terms of dominant words. The same situation is possible with groups related to the topic of “motivation”. A median value of maximum intersection between clusters and topics is 0.489, which suggests there is a strong interrelation between the obtained topics and group similarity in terms of users preference. That gives us confidence, that we are able to combine similarity graph and topic modelling in order to obtain interpretable picture of group preference in a scale of the whole social network.

5.4 Topics and Group Scale

At this stage, we perform an analysis of the dependency between the relative density of topic communities on each level (see details in Sect. 4.3). To present obtained results in an appropriate way, we divide topics according to levels, at which they reach peaks of relative density. As a result, we obtain 4 dominant levels (Fig. 7). Despite observing groups at level 1 and 10 in the distribution, the relative density equals to 0 for both of them because of the weak ties between groups in similar topics. Possibly it is an issue of data collection process, as we collect relatively large groups. From the other side, level 1 is poorly represented

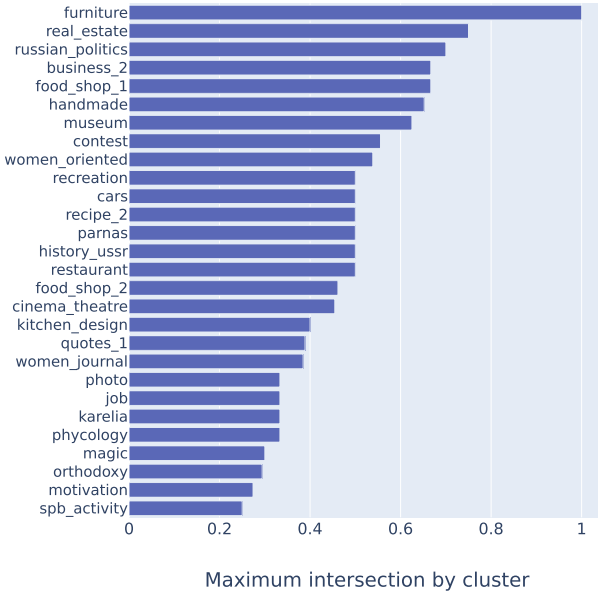


Fig. 6. Maximum intersection with cluster for each topic.

due to scale-free effect (instances are rare), level 10 is also poor. It is important to emphasize, that maximum possible value of relative density decreases with growing of dominant level, that should signify a tendency of small communities to be more specific, i.e. if one is interested in a certain topic, they subscribe for a different groups with higher probability.

Topics with the highest relative density at levels 2 and 3 (Fig. 7a and 7b) show similar patterns of having the highest density on rather small membership scale (up to 6k and 18k respectively), followed by decrease on the next levels, with a small possible rise afterwards. This could indicate that for the subset of topics, users tend to express more interest to less crowded groups in relation to bigger groups of the same topic. At the level 2 one can observe, that topic related to “Job” demonstrates smooth decrease with increasing level. Levels 3 and 4 have also strong preference, moreover, on level 6 there is another significant peak. Such results can be interpreted in the following way: small groups (level 2–3) represent jobs in a certain field, and users tend to subscribe them in order to find a job; for level 6 we have a peak possibly because there is some big groups-aggregators with a wide range of professions. Similar trends are observed for “Photo” and “Parnas” (district of Saint-Petersburg) at level 2 and for “Cinema and theatre”, “Recipe” at level 3. There is another interesting pattern: a topic starts with slightly low relative density and scatter relatively equally among other levels. This possibly means, that such kinds of topics do not tend to be specific and are relatively general at all levels: “Restaurant”, “Saint-Petersburg activity”, “Motivation”, or “Women journals” (except level 3 with high specificity).

In plots *c* and *d* of Fig. 7 the dominant topics at level 4 and 5 show the same patterns: relative density starts from the maximum value and then decreases among the next levels (“Kitchen design”, “Furniture”, “Karelia” at level 4 and “Orthodoxy”, “Magic” and “Real-estate” on 5) or equally scattered over the next levels (“Quotes”, “Psychology” on 4, “Women oriented” on 5). However, these two dominant levels bring us to a new pattern: some topics tend to show growth, the presence of one peak followed by a decrease. Appearance of such picture, especially if value of relative density on previous or next level is close, means, that such topics are popular among the social network, as a border between scale levels does not have a significant difference and there is a lot of groups with similar behaviour. In this case, the existence of a single peak become significant, as it is able to describe the tendency of a topic to be more general or specific, as it is able to describe the tendency of a topic to be more general or specific. In this way, for instance, for “Business” topic at level 4, density of the “growth” to the peak is prevalent, which suggests, that people tends to follow smaller groups more, because they are related to a more specific business. The same trend is observed for “Russian politics” at level 5. In contrast, for “Handmade” topic at level 5, the decrease in density over levels is prevalent, which means users tends subscribe to general groups, aggregating wider range of interesting things to make by own hands (which is more attractable for users).

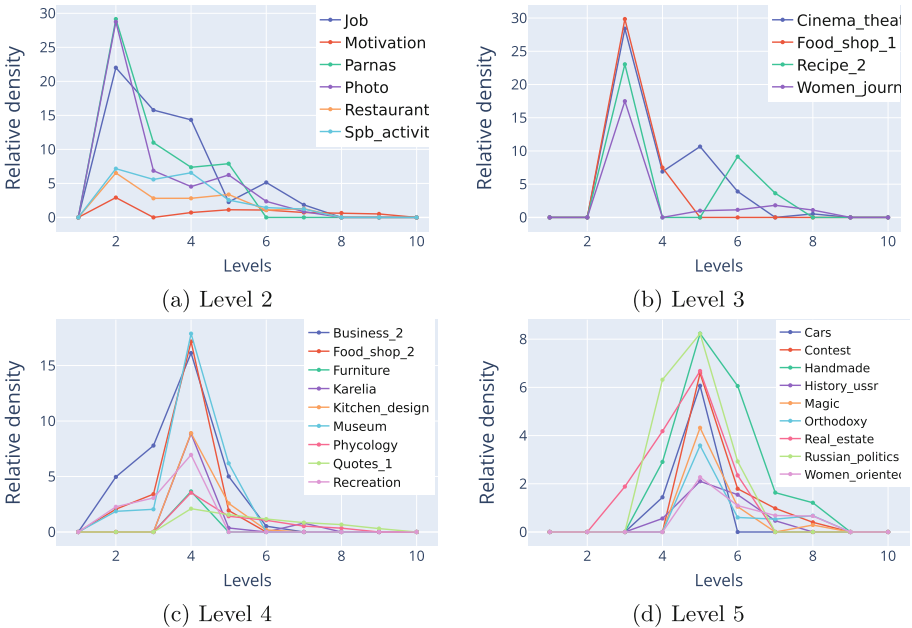


Fig. 7. Dominant topics on different levels

6 Conclusion and Future Work

Current paper concerns measuring the degree significance of interest topic in terms of their generality or specificity. To address this question, we divide all collected groups into 10 levels and measure the similarity between them on the based of the number of mutual followers. Then we perform topic modelling based on post texts in such groups. To obtain an interpretable picture of groups preference by users, we combine similarity graph with topic modelling and estimate their conformity: a median value of maximum intersection between clusters of similarity graph and topics is 0.489, which suggests a strong interrelation between the groups semantics and their coincidence in terms of users preference. Finally, we analyse the density inside the topic in relation to all other topics, at each considered level of group size.

Based on the analysis of the dependency between levels (in terms of group size) and relative density of topics, we conclude, that in general, relative density decreases with increasing of group size, which means small communities are more specific. Moreover, we uncover three patterns: 1) a topic appears at a certain level with a maximum relative density and followed by downfall on the next levels, i.e. this topic is clearly specific on the smallest level where it appears; 2) a topic starts with slightly low relative density and scattered equally among other levels: that a topic may be of general interest, but is divided in multiple subtopics, which become observable by a relatively high density on multiple levels; 3) topics tend to show growth, peak, and decrease of density, which is related to the popularity of the topic as a border between scale levels become neglected, but one can still distinguish the degree of specificity for them by analysis of the tendency density changes. All uncovered patterns interplay well with the semantics of topics.

Results give us a lot of perspectives for future studies, as we are able to characterise a specificity of a particular group and also characterise the level of topic involvement for a user or a local group. There is a great possibility to model interscale dependency between individuals, local groups of individuals, and the whole topics of interests. Moreover, obtained results give a good perspective to study and model topic interrelation and evolution. The complexity of other methods in this study is not higher than $O(N \cdot \log N)$, excluding only the proposed method of graph construction. It's computational complexity is $O(N^2)$, therefore one should elaborate on the more sophisticated approaches, for instance, machine learning techniques in the field of collaborative filtering algorithms, which allows for processing millions of groups, instead of thousands.

References

1. Backstrom, L., Huttenlocher, D., Kleinberg, J., Lan, X.: 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, pp. 44–54 (2006)
2. Butler, B.S., Bateman, P.J., Gray, P.H., Diamant, E.I.: An attraction-selection-attrition theory of online community size and resilience. *MIS Q.* **38**(3), 699–729 (2014)

3. Carr, P.B., Walton, G.M.: Cues of working together fuel intrinsic motivation. *J. Exp. Soc. Psychol.* **53**, 169–184 (2014)
4. Cheryan, S., Plaut, V.C.: Explaining underrepresentation: a theory of precluded interest. *Sex Roles* **63**(7–8), 475–488 (2010)
5. Cinelli, M., Brugnoli, E., Schmidt, A.L., Zollo, F., Quattrociocchi, W., Scala, A.: Selective exposure shapes the Facebook news diet. arXiv preprint [arXiv:1903.00699](https://arxiv.org/abs/1903.00699) (2019)
6. Derevitskii, I., Severiukhina, O., Bochenina, K.: Clustering interest graphs for customer segmentation problems. In: 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 321–327. IEEE (2019)
7. Dewey, J.: *Interest and Effort in Education*. Houghton Mifflin, Boston (1913)
8. Faralli, S., Stilo, G., Velardi, P.: Automatic acquisition of a taxonomy of microblogs users' interests. *J. Web Semant.* **45**, 23–40 (2017)
9. Gao, L., Wu, J., Zhou, C., Hu, Y.: Collaborative dynamic sparse topic regression with user profile evolution for item recommendation. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
10. Guleva, V., Vaganov, D., Voloshin, D., Bochenina, K.: Topology of thematic communities in online social networks: a comparative study. In: Shi, Y., et al. (eds.) ICCS 2018. LNCS, vol. 10860, pp. 260–273. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93698-7_20
11. Han, X., Wang, L., Crespi, N., Park, S., Cuevas, Á.: Alike people, alike interests? Inferring interest similarity in online social networks. *Decis. Support Syst.* **69**, 92–106 (2015)
12. Harackiewicz, J.M., Knogler, M.: Theory and application. In: *Handbook of Competence and Motivation: Theory and Application*, p. 334 (2017)
13. Hidi, S., Renninger, K.A.: The four-phase model of interest development. *Educ. Psychol.* **41**(2), 111–127 (2006)
14. Hulleman, C.S., Thoman, D.B., Dicke, A.-L., Harackiewicz, J.M.: The promotion and development of interest: the importance of perceived values. In: O'Keefe, P.A., Harackiewicz, J.M. (eds.) *The Science of Interest*, pp. 189–208. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55509-6_10
15. Hulleman, C., Dicke, A., Kosovich, J., Thoman, D.: The role of perceived social norms and parents' value in the development of interest in biology. In: Poster Presented at the Annual Meeting of the Society for Personality and Social Psychology, San Diego, CA (2016)
16. Ji, L., Liu, J.G., Hou, L., Guo, Q.: Identifying the role of common interests in online user trust formation. *PloS One* **10**(7), e0121105 (2015). <https://doi.org/10.1371/journal.pone.0121105>
17. Kapanipathi, P., Jain, P., Venkataramani, C., Sheth, A.: User interests identification on twitter using a hierarchical knowledge base. In: Presutti, V., d'Amato, C., Gandon, F., d'Aquin, M., Staab, S., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8465, pp. 99–113. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07443-6_8
18. McCaslin, M.: Co-regulation of student motivation and emergent identity. *Educ. Psychol.* **44**(2), 137–146 (2009)
19. McPherson, G.E., Osborne, M.S., Barrett, M.S., Davidson, J.W., Faulkner, R.: Motivation to study music in Australian schools: the impact of music learning, gender, and socio-economic status. *Res. Stud. Music Educ.* **37**(2), 141–160 (2015)

20. Newman, D., Lau, J.H., Grieser, K., Baldwin, T.: Automatic evaluation of topic coherence. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 100–108. Association for Computational Linguistics, Los Angeles (2010)
21. Newman, M.E.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**(23), 8577–8582 (2006)
22. Philippe, F.L., Vallerand, R.J., Lavigne, G.L.: Passion does make a difference in people's lives: a look at well-being in passionate and non-passionate individuals. *Appl. Psychol.: Health Well-Being* **1**(1), 3–22 (2009)
23. Piao, G., Breslin, J.G.: Inferring user interests for passive users on Twitter by leveraging followee biographies. In: Jose, J.M., et al. (eds.) *ECIR 2017. LNCS*, vol. 10193, pp. 122–133. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56608-5_10
24. Renninger, K.A., Hidi, S.: *The Power of Interest for Motivation and Engagement*. Routledge, Abingdon (2015)
25. Renninger, K.A., Hidi, S., Krapp, A., Renninger, A.: *The Role of Interest in Learning and Development*. Psychology Press (2014)
26. Rotgans, J.I., Schmidt, H.G.: The role of interest in learning: knowledge acquisition at the intersection of situational and individual interest. In: O'Keefe, P.A., Harackiewicz, J.M. (eds.) *The Science of Interest*, pp. 69–93. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55509-6_4
27. Schmutz, V., Stearns, E., Glennie, E.J.: Cultural capital formation in adolescence: high schools and the gender gap in arts activity participation. *Poetics* **57**, 27–39 (2016)
28. Segalovich, I.: A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. In: *MLMTA*, pp. 273–280. Citeseer (2003)
29. Song, Y., Elkahky, A.M., He, X.: Multi-rate deep learning for temporal recommendation. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 909–912 (2016)
30. Thoman, D.B., Sansone, C., Geerling, D.: The dynamic nature of interest: embedding interest within self-regulation. In: O'Keefe, P.A., Harackiewicz, J.M. (eds.) *The Science of Interest*, pp. 27–47. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55509-6_2
31. Traag, V.A., Waltman, L., van Eck, N.J.: From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**(1), 1–12 (2019)
32. Vaganov, D., Sheina, E., Bochenina, K.: A comparative study of social data similarity measures related to financial behavior. *Proc. Comput. Sci.* **136**, 274–283 (2018)
33. Vaganov, D.A., Guleva, V.Y., Bochenina, K.O.: Social media group structure and its goals: building an order. In: Aiello, L.M., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L.M. (eds.) *COMPLEX NETWORKS 2018. SCI*, vol. 813, pp. 473–483. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05414-4_38
34. Vorontsov, K.V.: Additive regularization for topic models of text collections. *Dokl. Math.* **89**(3), 301–304 (2014). <https://doi.org/10.1134/S1064562414020185>
35. Worsley, A., Wang, W., Ismail, S., Ridley, S.: Consumers' interest in learning about cooking: the influence of age, gender and education. *Int. J. Consum. Stud.* **38**(3), 258–264 (2014)
36. Zarrinkalam, F., Kahani, M., Bagheri, E.: User interest prediction over future unobserved topics on social networks. *Inf. Retrieval J.* **22**(1–2), 93–128 (2019)