Modelling Behavior Patterns in Cellular Networks

T. Couronne
Orange Labs
France Telecom R&D,
Paris, France
Email:Thomas.Couronne@orange-ftgroup.com

V. Kirzner
Institute of Evolution
University of Haifa
Haifa, Israel
Email:valery@research.haifa.ac.il

K. Korenblat, E.V. Ravve, Z. Volkovich
Software Engineering Department
Ort Braude College
Karmiel, Israel
Email:{katerina,cselena,vlvolkov}@braude.ac.il

***Abstract*—In this paper, we explore customer behavior in cellular networks. We develop a novel model of the fundamental user profiles. The study is based on investigation of activities of millions of customers of Orange, France. We propose a way of decomposition of the observed distributions according to certain external criteria. We analyze distribution of customers having the same number of calls during a fixed period. A segmentation of the population is provided by an approximation of the considered distribution by means of a mixture of several more "basic" distributions presenting the "granularity" of the user's activity. In order to examine the meaning of the found approximation, a clustering of the customers is provided using their daily activity, and a new clustering procedure is constructed. The optimal number of clusters turned out to be three. The approximation is the reduced in the optimal partition to a single-exponential one in one of the clusters and to two double-exponential in others. This fact confirms that the proposed partition corresponds to reliable consequential social groups.**

Keywords–Consumer behavior pattern; Market segmentation; Probability distribution; Mixture distribution model; Machine learning; Unsupervised classification; Clustering.

I. INTRODUCTION

General view of consumer behavior is a study how people, groups and companies purchase, work with and organize goods, services, ideas and knowledge in order to meet their needs and desires [1][2]. Such a multidisciplinary study strives to understand the decision-making processes of customers and serves as a basis for market segmentation. Through market segmentation, large mixed markets are partitioned into smaller sufficiently homogeneous sectors having similar needs, wants, or demand characteristics.

In the cellular networks context, the mentioned products and services can be expressed in spending of the networks resources such as the number of calls, SMS and bandwidth. In fact, market segmentation in this area is able to characterize behavior usage or preferences for each customers' sector; in other words, to typify the customers' profiles, aiming to use this pattern to intimately adopt specific products and services to the clients in each market segment.

The research, presented in this paper, is devoted to developing of a novel model of the fundamental user behavior patterns (user profiles) in the cellular networks. We base our study on analyzing of the underlying distribution of customers having the same number of calls during a fixed period, say a day. A segmentation of the population is provided by an approximation of the considered distribution by means of a mixture of several more "basic" distributions, which present the "granularity" of the user's activity. Actually, the mixture distribution models have come to be conventional in machine learning due to their fruitful applications in unsupervised classification (clustering), where the underling probability distribution is decomposed into a mixture of several simple ones, which correspond to subgroups (clusters) with high inner homogeneity.

Hypothetically, each one of these sets corresponds to a social group of users having its own dynamics of calls depending upon the individual group social parameters. As it will be demonstrated in this contribution, an empirical densities of the studied underlying distributions are monotone decreasing and do not exhibit multi-modality. These properties characterize mixtures of the exponential distribution [3][4]. Hence, in this research, an exponential distribution mixture model is applied, and a three-exponential distribution well-fits the needed target.

In fact, the common applications, for instance in clustering, of the known *Expectation Maximization algorithm*, which estimates parameters of mixture models, suggests the *Gaussian Mixture Model* of the data. However, many studies are recently devoted to analysis of non-Gaussian processes, which are often related to the power law distributions. Nevertheless, the very existence of such a law does not depend on the particular model, but rather it is a result of the process being non-Gaussian in its own nature. Such models arise in some fields of human endeavor. In fact, the Zipf's law declares that the words occurrences in a text collection is inversely proportional to its position in the sorted frequency list.

In order to explore the meaning of the found approximation, a clustering of the customers is provided using daily activity of the customers. Moreover, a new clustering procedure is constructed in the spirit of the bi-clustering methodology. The estimated optimal number of clusters turned out to be three; in addition, the mentioned approximation is the reduced in the optimal partition to a single-exponential one in one of the clusters and to two double-exponential in others. This fact confirms that the proposed partition corresponds to reliable consequential social groups. Here, we emphasize the fact that the similarity measure, applied in the clustering process, is formed without any reference to the previously discussed mixture model.

The results, reported in the paper, are obtained by means of a study of the daily activity of a real group of users during the period from March 31, 2009 through April 11, 2009. For each considered day, several million users in this group are active (making one or more calls), and the time location of each input or output call is known. The sets of active users on different days vary significantly.

 ISBN: 978-1-61208-513-5

The remainder of this paper is structured as follows. Section II is devoted to a distribution model of the user activity and its decomposition. Section III describes the customer clustering procedure and its evaluations. Section IV summarizes the paper.

II. DISTRIBUTION MODEL OF USER ACTIVITY

In this section, we consider a mixture model approximation of the underlying distribution of users having the same number of calls during a day (DSN distribution). We distinguish two types of user activity: input calls (Activity 1) and output calls (Activity 2). All users (about five millions) are divided into groups according to their number of calls per day. The i-th group contains all customers having exactly i calls per a day. The size of the i-th group is denoted by N_i.

Obviously, the groups' content and sizes are, generally speaking, not the same for different days. The amount groups with $i > 100$ is very small in the dataset. They are most likely containing "non-standard" users: sales agents, call centers and so on. We discard such groups together with users, who do not call at all in a given day. Actually, this lack of activity could be explained by factors, which are not directly related to the user activity on the network.

De facto, for all collected days, the curves are of almost the same monotonically decreasing form. On the other hand, it is naturally to assume that the underlying population is actually a mix of several different sub-populations. Practically, a mixture distribution model with exponential components appears to be an appropriate approximation to DSN. Mixture distribution models appear in many applications such as an inherent and straightforward tool in order to pattern the population heterogeneity. The assumption about exponential distributed mixture components commonly invokes in the study of lifetime or more universal duration data. We give the following simple k-finite exponential mixture model, having density function of the form

$$f(x) = \sum_{j=1}^{k} A_j exp(-t_j x), \qquad (1)$$

where A_j and t_j, $j = 1, ...k$ are non-negative numbers, and $\sum_{j=1}^{k} A_j = 1$.

For a given number of components k, the *Expectation–Maximization algorithm* is a traditional method for maximum likelihood estimates for finite mixtures. This well understandable technique is much admired because it satisfies a monotonic convergence property and can be easily implemented. Nevertheless, there are several known drawbacks of the method. In fact, if there are multiple maxima, the algorithm may discover a local maximum, which is not a global one. Moreover, the obtained solution strongly depends on the initial values selection (see, e.g. [5]).

In this contribution, another approach in the spirit of the linear regression methodology is applied without any prior suggestion about the components number k. For this purpose, we initially form the explanatory variable $X = (1, 2, ..., 100)$ and the response Y, which for each value $x \in X$ is composed of the logarithm values of the normalized frequencies of DSN in a day: $ln\,(f(x))$.

Using the standard simple regression methodology (see, e.g. [6]), a linear regression model is identified: $Y = a + bX$ and the first estimation of the density $f(x)$ in (1) is constructed $f^{(1)}(x) = A_1 \exp(-t_1 x)$, for $A_1 = \exp(a)$ and $t_1 = -b$. In the next step, a new response is built $Y = ln\left(f(x) - f^{(1)}(x) + C\right)$, where C is a sufficiently positive number, insuring that $f(x) - f^{(1)}(x) + C > 0$ for all x and j; then, the described procedure is repeated and so on. In each step, p-value coefficient of significance:

$$F = \frac{R^2(X,Y)}{1 - R^2(X,Y)}(100 - 1) \qquad (2)$$

is calculated. The process is stopped if the actual p-value is greater than the traditional level of significance 0.05. Here, $R(X,Y)$ is the *Pearson correlation coefficient* between X and Y. For all cases of daily activity, the method has been stopped after three components were extracted.

TABLE I. p-VALUES

component number	1	2	3	4
p-value	0	8.6e-06	0.025	0.282

The parameters of model (1), calculated for each of the 13 studied days, demonstrate high stability of the exponent indexes (t_1, t_2, t_3), which are practically independent on time but are rather somewhat different on the weekends, i.e. Saturday (4.04 and 11.04) and Sunday (5.04). Amplitudes A_1, A_2, A_3 differ to a greater degree (in percentage terms). Thus, the absolute number of active users varies from day to day to a greater extent than the distribution pattern, which actually corresponds to a set of exponent indexes. The p-values, calculated for the first of the considered days, are presented in Table I.

In the case of input calls, the ratio of the exponent indexes is: $3 \cdot t_1 \approx t_2, 3 \cdot t_2 \approx t_3$. In the case of output calls, this ratio is somewhat different: $2 \cdot t_1 \approx t_2, 3.5 \cdot t_2 \approx t_3$. The decay value, x_0, of each component in (1) is chosen to normalize the component value at this point to 1. The components are not equivalent in the sense of their decay value. Thereby, the exponent with index $t_3 = 1.0$ and amplitude $A_3 = 500,000$ (these parameter values are typical of one of the three exponents, which constitute the daily activity) already decays at $x_0 = 13$. For the second typical pair of parameter values ($t_2 = 0.33$ and $A_2 = 400,000$), the decay occurs at $x_0 = 39$. The exponent with $t_1 = 0.12$ and $A_1 = 90,000$ has the longest effect on DSN ($x_0 = 95$).

Accordingly, two of the three components that describe user activity disappear in the middle of the considered interval of calls. Only the third exponent continues, and its values can be considered to represent the "asymptotic behavior" of the distribution. The relatively complex nature of the obtained empirical distribution model of user activity may be indicative of the heterogeneity of the entire set of users. This set is conceivably composed of a few groups such that the total user activity in a group is described by a certain simpler distribution.

Obviously, the social status, gender and age of the users affect their activity on telephone networks; however such type of personal data is not available for us. Therefore, in the following section we divide the users into groups based merely on the features of their individual activity during a given day.

 ISBN: 978-1-61208-513-5

III. USER CLASSIFICATION

As it was justified in the previous section, we proceeded from the assumption that the obtained three-component exponential mixture model reflects the inner customers' behavior patterns, demonstrated by the data. In order to identify these patterns, all the users under investigation are divided into groups according to a comparable daily performance.

A straightforward clustering of the original data is hardly expected to deliver a robust and meaningful partition. Actually, such a situation is a common place in the current practice. Moreover, in many applications, the aim is to reveal not merely potential clusters, but also a quite small number of variables, which adequately settle that partition. For instance, the sparse K-means, proposed in [7], at once discovers the clusters and the key clustering variables.

A procedure in the spirit of such a bi-clustering methodology, where features and items are simultaneously clustered, is applied in this paper. First of all, 24 hours inside a day (the features) are clustered according to the corresponding users' activity. In the next step, the users are divided in groups according to their occurrences in the hour's partition. As a result, a sufficiently robust clustering of users is obtained together with the clusters' description in terms of the call activity.

A. Clustering of hours

In order to outline a similarity between hours in a day, we consider each hour as a distribution of users across the actual numbers of calls within this hour. It means: how many people did not call at all in this hour, how many people called just one time, two times and so on.

A dissimilarity between hours from the point of view of the users' behavior can be naturally characterized by a distance between the corresponding distributions. Generally speaking, any asymptotically distribution-free statistic is suitable for this purpose. In this study, we employ the well-known *Kolmogorov-Smirnov (KS) two sample test statistic* (see, e.g., [8][9]), which is actually the maximal distance between two empirical normalized cumulative distribution functions.

Calculating the KS-distance for each pair of hours, we get a 24×24 distance matrix. Now, the *Partitioning Around Medoids (PAM) clustering algorithm* (see, e.g., [10]) is applied in order to cluster the data. This algorithm operates merely with a distance matrix, but not with the items themselves; it is feasible for small data sets (such as considered one composed from 24 hours) and a small number of clusters. In order to divide a data set into k clusters using PAM, firstly, k objects from the data are chosen as initial cluster centers (medoids) with the intention to attain the minimal total scattering around them (to reduce the loss function value). Then, the process iteratively replaces each one of these center points by non-center ones with the same purpose. If any further change cannot improve the value of the loss function then the procedure ends.

Except of the clustered data, PAM includes as an input parameter the number of clusters k. Hence, the first step of our procedure is devoted to estimation of the optimal number of hour's clusters. For this purpose, the well-known *Silhouette coefficient* of [11] is employed. Here, ideas of both cohesion and separation are combined, but for individual points, as well as for partitions. For each point, the Silhouette index takes values in $[-1, 1]$ interval, such that the Silhouette mean value, calculated across the whole data, close to one specifies "well clustered" data, and value -1 characterizes a very "poor" clustering solution. Therefore, the Silhouette mean value, found for several different numbers of clusters, can indicate the most appropriate number of clusters by its maximal value. The number of clusters was checked in the interval of $[2-10]$, and the optimal one was found to be 3 for all the considered data sets (i.e., for all considered days). An example of Silhouette plots (for 05.04 and 10.04) is shown in Fig. 1.

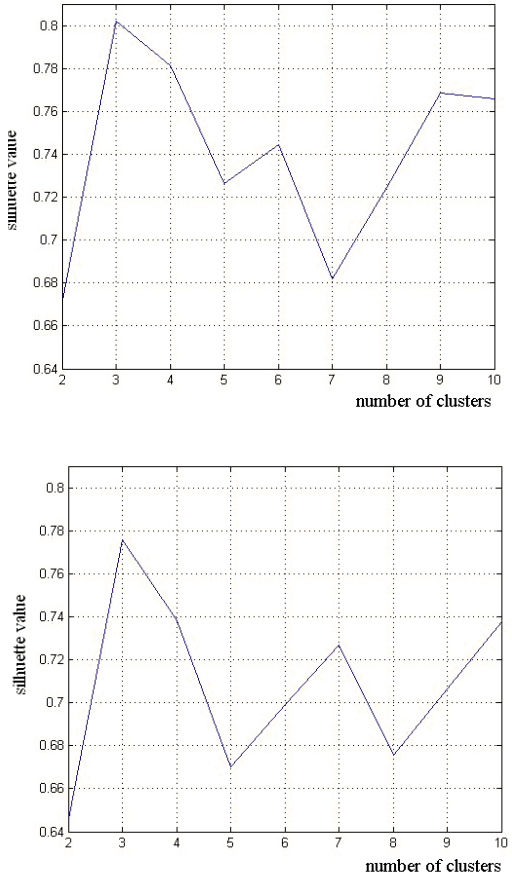

Figure 1. Silhouette plots for 05.04 (upper) and 10.04 (lower)

From the observed partition of 24 hours into 3 hour clusters, it can be concluded that although the partitions slightly depend on the particular data set (date), the overall structure of the clusters is preserved. Namely, there is a silent 'night' cluster, an active 'day' cluster, and a 'morning/evening' cluster.

1) Clusterization procedure: Now, every user is represented by means of a three dimensional vector (r_1, r_2, r_3), where r_i is the ratio of a user's activity during a cluster of hours number i. More precisely, it is a fraction of a user's calls during the cluster i in the total number of calls during a day. The proposed resampling clustering procedure is based on the well-known K-*means* (see, e.g., [12]) algorithm, implementing de-facto the idea, proposed in [13].

The K-means algorithm has two input parameters: the

number of clusters k and the data set to be clustered X. It strives to find a partition $\pi(X) = \{\pi_1(X), \ldots, \pi_k(X)\}$ minimizing the following loss function

$$\rho_{\{c_1,\ldots c_k\}}(\pi(X)) = \frac{1}{N}\sum_{j=1}^{k}\sum_{x\in\pi_j(X)} \|x - c_j\|^2, \qquad (3)$$

where c_j, $j = 1, ..., k$ is the mean position (the cluster centroid) of the objects belonging to cluster $\pi_j(X)$, and N is the size of X. Initially, the centroid set can be predefined or chosen randomly. Using the current centroid set, the K-means algorithm assigns each point to the nearest centroid, aiming to form the current clusters, and, then, recalculates centroids as the clusters means.

The process is reiterated until the centroids are stabilized. In the general case, as a result of this procedure, the objective function (3) reaches its local minimum. As a matter of fact, in the K-means algorithm, a partition is unambiguously defined by the centroid set and vise versa. Moreover, in the general case, the loss function (3) can be used for assessing the quality of arbitrary partition $\widehat{\pi}(X)$ with respect to the given set of centroids $\{c_1, \ldots c_k\}$.

The resampling procedure allows partitioning a large data set, based upon partitioning its parts. The algorithm is presented below:

Algorithm 1: **Input:**

- X - dataset to be clustered;
- k - the number of clusters;
- N - the number of samples;
- m - the sample size.
- ε - the threshold value.

Procedure:

1) Randomly draw N samples of size m from X without replacement.
2) For each sample S_i
   a) In the first iteration, the centroid set C is chosen randomly.
   b) Clustering S_i by K-means algorithm with starting from the given centroid set C.
   c) Clustering $\pi(X)$ of the whole data set by assignment to the nearest centroid using centroids obtained in the previous step.
   d) Calculate the object function value of $\pi(X)$ according to (3).
3) Choose from a set $\{S_1, \ldots, S_N\}$ a sample S_0 with the minimal object function value.
4) If the first iteration is being processed or if the absolute difference between two minimal object function values calculated for two sequential iterations is greater than ε, replace C with the set of centroids of $\pi(S_0)$, and return to step 2; otherwise stop.

2) Choosing number of users' clusters: In order to evaluate the optimal number of clusters, it is natural to compare stability of the obtained partition for different cluster numbers. To this end, we repeat the user clustering procedure ten times on the same data set and evaluate the Rand index value between all obtained partitions. The *Rand index* [14], represents the measure of similarity between two partitions. It is calculated by counting the pairs of samples, which are assigned to the same or to different clusters in these partitions. The closeness of the Rand index value to 1 indicates similarity of the considered partitions.

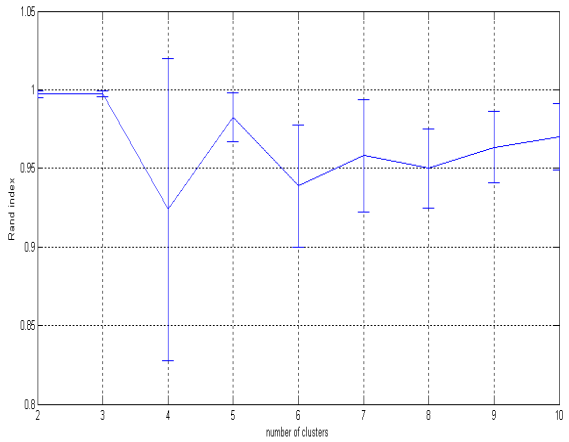

Figure 2. Rand index plot for the dataset 01.04

TABLE II. MINIMUM OF AVERAGE DISTANCES TO THE NEAREST CENTROID FOR THE FIRST 5 ITERATIONS OF RESAMPLING PROCEDURE

iteration num	1	2	3	4	5
min avg of dist	0.014487	0.013302	0.013295	0.013309	0.0132901

For the same purpose also *Adjusted Rand index* [15], which is the corrected-for-chance version of Rand index, can be used. However, in our consideration, it is suitable to use the regular one because it well reflects partitions' closeness. The mean value of the obtained Rand indexes naturally characterizes partition stability by its maximal value. Thereby, the 'true' number of clusters corresponds to the most stable partition.

B. Experimental study

1) 'True' number of clusters estimation: In order to estimate the optimal number of clusters in the users' clusterization procedure, we repeat the clustering stability evaluation procedure, described in Section III-A2, for each of the possible numbers of clusters in the interval $[2, 10]$. The results for all dates are very similar. Fig. 2 demonstrates an example of Rand-index curve for 01.04. It is easy to see that the maximal stability attitudes appear for $N = 2$ and $N = 3$.

Recall that the purpose of the user clustering is to recognize behavior patterns, which represent the general structure of the user population. Let us consider two possible estimators for 'true' number of clusters from this point of view. We describe a behavior pattern via an average level of the users' activity within each of 3 hour clusters, defined in Section III-A. In this way, we take a three-dimensional representation of users and calculate the mean as well as standard deviation of each coordinate in each user cluster.

The user activity patterns, found for 01.04, are shown in Fig. 3 by means of the error bar plot of values in each hour cluster. Recall that for the given data we obtained a 'night'

 ISBN: 978-1-61208-513-5

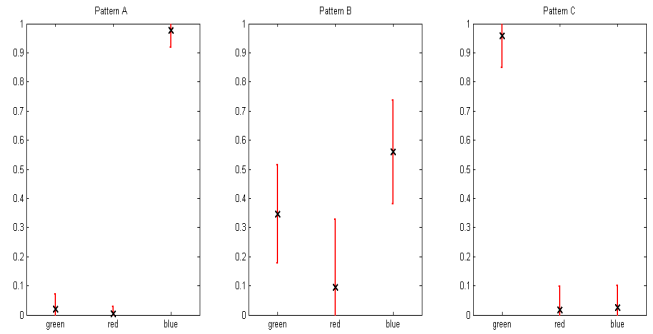

Figure 3. Profiles of 3 customer clusters for work day (01.04; Activity 1)

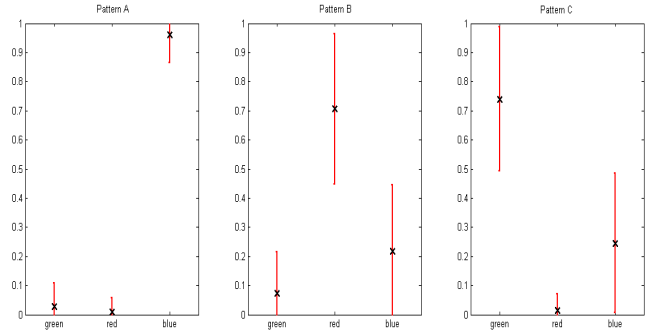

Figure 4. Profiles of 3 customer clusters for off day (05.04; Activity 1)

cluster with hours 1-8; a 'day' cluster with hours 11-21; and a 'morning/evening' cluster containing hours 9-10 and 22-24. For example, pattern A (the left panel in the picture) is characterized by the prevalence of the day activity since the average activity value is 0.84 for the 'day' hour cluster, in comparison with the values of 0.09 and 0.06 for the other hour clusters. Similarly, the behavior pattern B (the middle panel) describes users with significant activity in all hour clusters, while the pattern C (the right panel) is characterized by high activity in the morning-evening hours.

The obtained result shows that we have a "clear" partition into 2 clusters and that one of them is well divided into 2 more sub-clusters. In fact, the two-clusters partitions contain the cluster corresponding to Pattern B and the united cluster for Patterns A and C. For our purposes, therefore, it is natural to choose 3 as the "true" number of clusters. Actually, it is a common situation in cluster analysis, where the "ill-pose" number of clusters determination task can have several solutions depending on the model resolution.

2) Procedure convergence: Now, we demonstrate that the resampling clustering procedure converges very fast. Table II shows the minimal objective function values for the first five iterations of the resampling procedure, conducted on 100 samples for $k = 3$ (for others k the situation is similar). The results show that even in the second iteration the minimal average of the distances does not change significantly as compared to the first iteration. In the subsequent iterations, this value remains constant to within 0.0001.

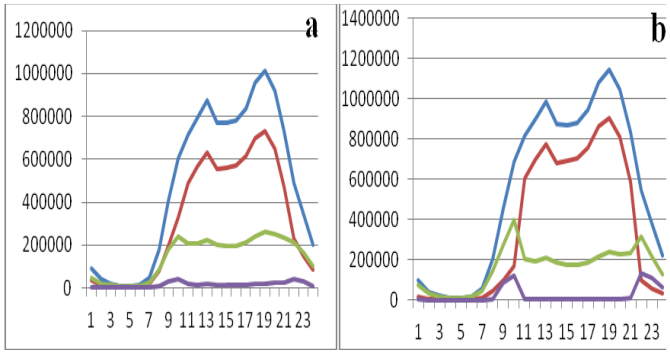

Figure 5. (a) Distribution of Activity 1 for the clusters obtained for Activity 2. (b) Distribution of Activity 1 for the clusters obtained for Activity 1. Date: 08.04.

3) Profile stability: Further, we use the behavior patterns for comparison of the results of our procedure on different datasets. The obtained results show that they are stable both for work days and off days. However, the difference between work and off days is significant (see Fig. 3 and 4 for comparison).

Although, qualitative descriptions of profiles are very similar in both cases: pattern A with prevalent "day" activity; pattern B with significant activity throughout 24 hours and pattern C with prevalent "morning-evening" activity; in off days higher "night" activity is detected.

C. Call activity, associated within patterns

Let us consider the call activity of users, located in each one of the found clusters. The total activity of all the users within a day has a density with two peaks. One of them is placed in the workday middle, and the second one, the higher peak, is located in the period after 7 p.m such that a local activity minimum is observed immediately after. The shape of the corresponding density in the first cluster (A) is actually the same. However, the user's activity almost does not vary in the second cluster (B), i.e. the density curve has several insignificant peaks, and the activity decreases at 10 p.m. The total activity of the users belonging to cluster three (C) has two peaks located in the morning and in the evening of a day.

Furthermore, we observe that the distribution of calls during a day for all three clusters is almost independent on the activity type, see Fig. 5. Here, the blue curves corresponds to the total activity densities of all the users; the red, green and brown ones give the total activity densities for clusters 1, 2 and 3, respectively. Note that both activity types have the same distribution shapes.

1) Features of the cluster model parameters: The model, which we use, reveals major differences between the DSN of the entire set of users and the DSN's for the individual clusters. For Activity 1, the DSN for Cluster 1 is almost always best fitted by a single exponent. However, in more than half of the cases, the DSN for Cluster 2 is fitted by two exponents. Moreover, during the weekend period, the curve is fitted by three exponents. The DSN for Cluster 3 is usually fitted by two exponents, while the three-exponent fit sometimes arises without regard for the day of the week. For Activity 2, the above regularities are more pronounced for Clusters 1 and 2, since all the best fits for Cluster 3 are two-exponential.

 ISBN: 978-1-61208-513-5

Our results demonstrate an obvious simplification of the DSNs for Clusters 1-3 as compared to the DSN for the total set of users. Nevertheless, joining any two of these clusters results in a three-component DSN. At the same time, random partition into three clusters (with the same number of users as in the calculation of Clusters 1, 2, 3 as mentioned above) yields the same three exponent indexes, $t_1 = 0.11$, $t_2 = 0.31$ and $t_3 = 1.01$ for all three clusters. The results coincide with those calculated for the total set of users on the same day.

Thus, simplification of the cluster model shows that the partition into Clusters 1-3 actually reflects different activity characteristics for different groups of users. There are some differences on the weekends, but on the whole the parameters of a particular DSN are the same for each day. Note also that the DSN's of Clusters 2 and 3 are not in the least close to the second or third component (exponent) of the total set DSN. Indeed, in our model, the DSN of Cluster 2 consists mainly of two exponents, with one exponent disappearing at the decay value of 30, while the other as a rule not decaying up to the value of 70. The DSN of Cluster 3 also has long-lasting components (up to 100 and more).

IV. Conclusion

In the present study, we are interested in the mechanisms, which generate non-Gaussian distributions. We investigate the reason that non-Gaussian distributions occur in the social sciences. Internet activity and, in particular user activity on social networks, appears to be an appropriate area for such analysis. Numerous studies suggest different models of social networks and try to link particular network characteristics to some measure of the user activity. These characteristics often obey the hyperbolic law in one form or another.

Although, the social activity distribution of a population takes a specific and constant form, it can be assumed that the observed distribution is in some sense an averaged one. Obviously, it is composed of various types of distributions, generated by different social layers. We have in mind not only the groups, arising from the simplest types of differences such as age and gender, but also the more complex features of the population under consideration. It can be assumed that the demonstration of the hyperbolic law or, in contrast, the combination of distribution laws for various social groups, depends on the nature of the user joint activity. In some cases, each user's action is in some sense sequential, so that their average behavior can be considered in the framework of a single law.

An example of parallel user activity is the number of records in an email address book, cf. [16]. In cases, where users' actions occur in parallel, each user group, which is uniform with respect to some criterion, can generate its own law of activity distribution. Since telephone calls are also more likely to be a parallel user's activity in the sense.

In this research, we expected to find that the observed distribution of calls is the sum of several distribution functions, corresponding to different social groups of users. The limited number of these groups is an important prerequisite for such differentiation because averaging over the groups is absent in this case. In [17], we introduced the notion of user strategy (with respect to alternating different types of telephone activity) and showed that the number of different strategies is small.

Therefore, we expected to obtain a small number of groups with equivalent user activity. Having no real-life socio-relevant parameters, we assumed that the peculiarities of a user's activity during a day may correlate with the user's social status. Finally, we partitioned the results into three clusters, with 70, 21, and 9 user percentages in these clusters. We showed that these clusters have simpler distribution functions than those for the total population.

References

[1] I. Simonson, Z. Carmon, R. Dhar, A. Drolet, and S. Nowlis, "Consumer research: in search of identity," Annual Review of Psychology, vol. 52, 2001, pp. 249–275.

[2] P. Kotler and K. Keller, Marketing Management, ser. MARKETING MANAGEMENT. Pearson Prentice Hall, 2006.

[3] N. Jewell, "Mixtures of exponential distributions," The Annals of Statistics, vol. 10, no. 2, 1982, pp. 479–848.

[4] J. Heckman, R. Robb, and J. Walker, "Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments," Journal of the American Statistical Association, vol. 85, no. 410, 1990, pp. 582–589.

[5] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, vol. 39, no. 1, 1977, pp. 1–38.

[6] J. Kenney and E. Keeping, "Linear regression and correlation," in Mathematics of Statistics: Part 1, 3rd ed. NJ: Princeton, Van Nostrand, 1962, ch. 15, pp. 252–285.

[7] D. Witten and R. Tibshirani, "A framework for feature selection in clustering," Journal of the American Statistical Association, vol. 105, no. 490, 2010, pp. 713–726.

[8] R. Lopes, P. Hobson, and I. Reid, "The two-dimensional Kolmogorov-Smirnov test," in Proceeding of the XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Nikhef, Amsterdam, the Netherlands, April 23-27, 2007. Proceedings of Science, 2007.

[9] ——, "Computationally efficient algorithms for the two-dimensional Kolmogorov-Smirnov test," Journal of Physics: Conference Series, vol. 119, no. 4, 2008, p. 042019.

[10] L. Kaufman and P. Rousseeuw, Finding groups in data: an introduction to cluster analysis, ser. Wiley series in probability and mathematical statistics. New York: Wiley, 1990, a Wiley-Interscience publication.

[11] P. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, 1987, pp. 53 – 65.

[12] D. MacKay, Information Theory, Inference & Learning Algorithms. New York, NY, USA: Cambridge University Press, 2002.

[13] J. Kogan, C. Nicholas, and M. Teboulle, Grouping Multidimensional Data: Recent Advances in Clustering, 1st ed. Springer Publishing Company, Incorporated, 2010.

[14] N. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Is a correction for chance necessary?" in Proceedings of the 26th Annual International Conference on Machine Learning, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 1073–1080.

[15] S. Wagner and D. Wagner, "Comparing Clusterings – An Overview," Universität Karlsruhe (TH), Tech. Rep. 2006-04, 2007.

[16] M. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," Physical Review E, vol. 66, Sep 2002, p. 035101.

[17] T. Couronné, V. Kirzhner, K. Korenblat, and Z. Volkovich, "Some features of the users activities in the mobile telephone network," Journal of Pattern Recognition Research, vol. 1, 2013, pp. 59–65.

 ISBN: 978-1-61208-513-5