

## Формирование кластеров в сложной сети, отражающей процессы кооперации в социальных группах

В. О. Недайвозов,<sup>\*</sup> В. В. Макаров,<sup>†</sup> М. В. Горемыко,<sup>‡</sup> Д. В. Кирсанов<sup>§</sup>  
<sup>1</sup>Саратовский государственный технический университет имени Гагарина Ю.А.  
Россия, 410056, Саратов, ул. Политехническая, 77  
(Статья поступила 23.07.2017; Подписана в печать 20.09.2017)

Изучены процессы кластеризации в многослойных социальных сетях коллабораций и профессионального взаимодействия, выявлены основные принципы, отвечающие за взаимосвязь топологии и динамических характеристик в подобных сетях.

PACS: 05.45.-a.

УДК: 519.179.2:51.77

Ключевые слова: теория сложных сетей, социальная сеть, осциллятор, нелинейные колебания.

### ВВЕДЕНИЕ

Сетевые структуры встречаются во всех сферах человеческой деятельности [1–3]. Их изучение позволило лучше понять множество сложных явлений — распространение информации в современном обществе [4], передачу болезней [5], функционирование сложных биологических процессов [6], а также способствовало развитию методов оптимизации многих искусственных систем [7]. Одной из ключевых является сфера социального взаимодействия [8–11], как процесса воздействия людей и групп людей друг на друга. В целом все явления такого типа сводятся к кооперации, как сотрудничеству для решения общих задач, или же конкуренции — борьбе людей или групп людей за обладание какими-либо благами или возможностями. Анализ группового социального взаимодействия является актуальной задачей и его понимание позволяет правильно интерпретировать глобальные процессы, происходящие в современном обществе [12].

Основной особенностью такого взаимодействия является то, что существует очень много способов его осуществления между одними и теми же людьми: один человек может иметь личные, профессиональные и другие типы связей в реальном мире, но также может являться владельцем нескольких учетных записей в так называемых онлайн-системах. То, что люди имеют дело с этим многообразием типов социальных связей, подразумевает, что все эти вышеупомянутые отношения проецируются в один уровень, но на самом деле, не все реальные процессы могут быть смоделированы на такой упрощенной сети контактов. Например, в сообществе музыкантов распространение новых трендов и течений может происходить со значительным участием интернет-технологий, в то время как запись совместного произведения очень часто предполагает непосредственное физическое взаимодействие между исполнителями.

Анализ структур, схожих с исследуемой в данной работе сетью профессионального взаимодействия, интересен с нескольких точек зрения. Во-первых, на его основе можно лучше понять принципы распространения различных направлений и новых течений в данной среде, во-вторых, выявленные физические закономерности позволяют строить новые алгоритмы генерации сетей, имеющих схожую с реальными системами структуру, а также создавать экспертные системы, позволяющие такие структуры выявлять. Очевидно, что исследования такого типа требуют тщательного выбора принципов построения сетевой модели, содержащей информацию о реальной системе.

### 1. НАБОРЫ ДАННЫХ

В данной работе проведен анализ профессиональных отношений в социуме на основе имеющихся статистических данных о мировом сообществе музыкантов, как частном примере социального взаимодействия. Данные для исследования были получены из базы музыкальных данных Allmusic [13] — достоверной, моделируемой профессиональными редакторами платформе, и представляют собой два набора параметров, где первый отражает жанровую принадлежность исполнителей, а второй — наличие, если таковое имелось, коллабораций между ними. Два элемента первого набора данных были соединены связью, если по мнению экспертов они играли в одном музыкальном стиле. Во втором наборе два исполнителя получали связь, если они работали над одним или более альбомов совместно. Первоначально, первый набор данных о жанровой принадлежности артистов содержал 32 377 вершины и 117 621 связь, и второй — о коллаборациях между ними — 34 724 вершины и 123 082 связи. В целом у этих матриц было 8 509 общих вершин, но некоторые из них имели связи лишь в пределах одного слоя, а на другом не имели их вообще. Удалив из наборов такие узлы, была выделена главная связанная компонента сети [14], с которой проводилась дальнейшая работа. Такой подход является типичным

\*E-mail: fdf\_het@mail.ru

†E-mail: gormv67@mail.ru

‡E-mail: vladmak404@gmail.com

§E-mail: dankirsdot@gmail.com

Таблица I: Выделение главной компоненты сети

	Сеть жанровой принадлежности			Сеть коллабораций		
	Исходные данные	Пересечение	Главная компонента	Исходные данные	Пересечение	Главная компонента
Узлы	32 377	8 509	4 673	34 724	8 509	4 673
Связи	117 621	24 950	16 150	123 122	20 232	15845

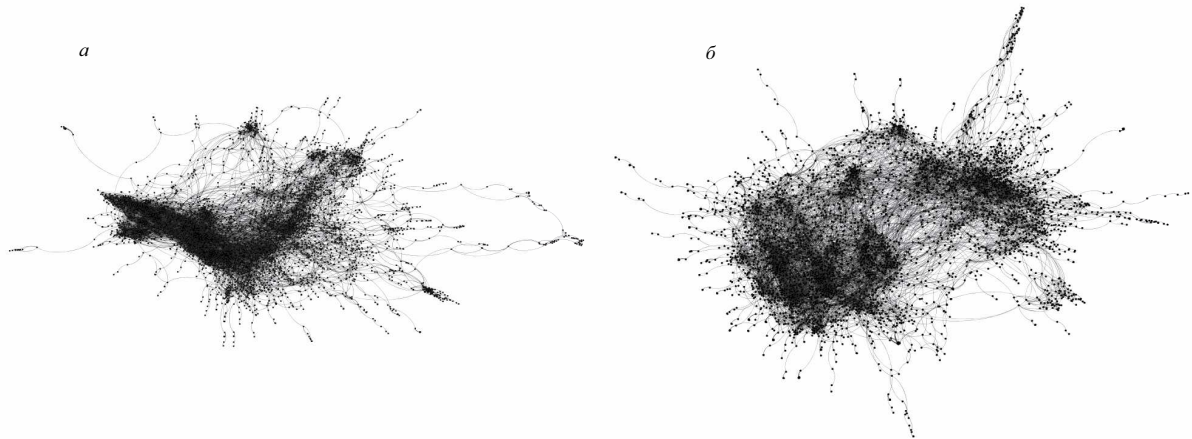


Рис. 1: Визуализация исследуемой сети профессионального взаимодействия: *a* — сеть жанровой принадлежности, *б* — сеть коллабораций

при исследовании реальных сетевых структур (в том числе и социальных) и позволяет избежать искажений характеристик исследуемой системы, вносимых большим количеством узлов, не имеющих связей. Сводная информация о количестве музыкантов и связей между ними представлена в табл. 1.

## 2. МОДЕЛЬНАЯ СИСТЕМА

Исследуемая система представляет собой многослойную сеть (multiplex network) элементов, каждый из слоев которой содержит в себе  $N = 4673$  элемента (музыканта), где первый слой построен на основе данных о связи исполнителей согласно их жанровой принадлежности, второй — на основе коллабораций между ними. Визуализация модельной сети представлена на рис. 1.

Видно, что на каждом уровне есть как узлы, обладающие большим количеством связей и являющиеся так называемыми хабами (hubs), так и множество узлов, находящихся на периферии сети и не являющиеся очень значимыми для данной структуры. В то же время, представленные на рис. 2 степенные распределения слоев показывают, что и первый слой, построенный

по принципу соединения узлов сети согласно жанровой принадлежности музыкантов, так и второй, определяемый коллаборациями, обладают свойством свободной масштабируемости [15], которое присуще всем реальным сетям. Такой результат является важным для представленной модели, так как он подтверждает верность ее построения.

Чтобы оценить структурные характеристики исследуемой сети были рассчитаны три её параметра. Первый из них — локальный коэффициент кластеризации, введенный в работе [16], который характеризует степень взаимодействия между собой ближайших соседей узла  $i$ .

Коэффициент кластеризации данного узла равен вероятности того, что два ближайших соседа узла являются ближайшими соседями друг для друга. Иными словами, если узел  $i$  имеет  $q_i$  ближайших соседей с числом  $t_i$  связей между ними, то локальный коэффициент кластеризации рассчитывается согласно следующей формуле:

$$C_i = \frac{t_i}{q_i(q_i - 1)/2}. \quad (1)$$

Число  $t_i$  есть суммарное число треугольников — циклов длины 3 — прикрепленных к узлу  $i$ ,

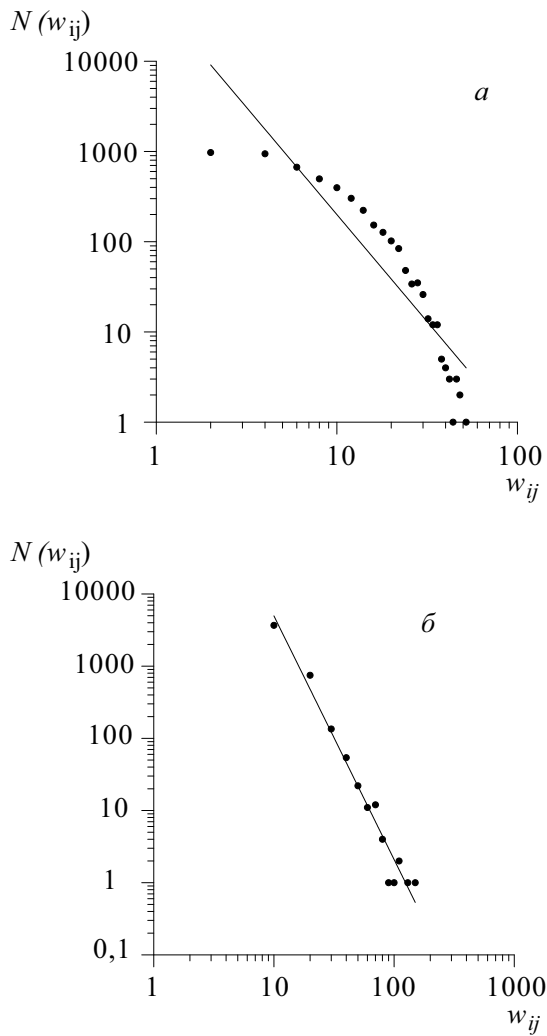


Рис. 2: Степенное распределение весов связей в структурах: а — жанровой принадлежности, б — коллабораций

а  $q_i(q_i - 1)/2$  — максимальное количество треугольников. Если все ближайшие соседи узла  $i$  взаимосвязаны, то  $C_i = 1$ , в противном случае, когда между соседними для узла  $i$  элементами нет подобной связи, как например в структуре сети, подобной иерархическому дереву,  $C_i = 0$ .

Второй параметр — центральность собственного вектора (eigenvector centrality, EVC) была определена в начале семидесятых годов [17]. Основной её принцип применительно к сетям социального характера состоит в том, что не так важно, с каким количеством людей поддерживает контакт индивид, но важно, насколько важны, т.е. центральны, в свою очередь эти люди. Это определяет важность самого индивида. Таким образом, очевидно, возникает рекурсивное определение центральности собственного вектора: важность (центральность) индивида зависит от важности его соседей, которая, в свою очередь, зависит от его важности. Смысл такого рекурсивного определения заключается в том, что оно позволяет определять ту небольшую

группу узлов, которая на самом деле имеет заметное влияние в масштабе всей сети. Если пытаться определить важность узла только лишь исходя из числа его соседей, имеется риск появления ситуации, в которой важнейшими узлами сети будут названы центры изолированных кластеров. Но, снова возвращаясь к социальным структурам, такие центры будут иметь влияние в очень ограниченном масштабе, так как их влияние не выйдет за пределы их непосредственных соседей. Таким образом — собственный вектор центральности, характеризует меру влияния узла в сети. Каждый узел получает относительную оценку важности  $E_i$  в соответствии с концепцией, что соединение узла  $i$  с узлом  $j$ , имеющим большую центральность  $E_j$ , внесет больший вклад в оценку рассматриваемого узла  $i$ , чем аналогичные соединения с узлами, имеющими малую центральность:

$$E_i = \frac{1}{\lambda} \sum_{j \in M} w_{ij} E_j. \tag{2}$$

Третьим параметром, также играющим важную роль для процессов, протекающих в реальных сетях, является количество кратчайших путей  $B$ . Когда возникает необходимость сообщения между двумя удаленными узлами сети, кратчайший путь обеспечивает оптимальное решение данной задачи, и именно по этой причине число кратчайших путей, проходящих через узел, является важной величиной для характеристики внутренней структуры сети.

На рис. 3 представлены зависимости рассчитанных характеристик друг от друга. Как видно, центральность по собственному вектору находится в обратной корреляции с коэффициентом кластеризации (рис. 3,а,б), что говорит о существовании в сети определенного количества кластеров, связанных между собой через ограниченный набор узлов. Видно, что существуют узлы с большим значением кратчайших путей, проходящих через них, но характеризуются очень малым коэффициентом кластеризации. Это позволяет сделать вывод, что через такие узлы являются перешейками между большими кластерами. В то же время эти связующие узлы обладают достаточно небольшой центральностью (рис. 3,в,г), что, в связи со специфичностью этой характеристики, подтверждает нахождение этих узлов вне существующего основного кластера, который, очевидно, и агрегирует большие значения центральности. Интересным моментом также является тот факт, что несмотря на обратный характер корреляции двух представленных выше зависимостей, центральность по собственному вектору демонстрирует схожую зависимость с коэффициентом кластеризации. Это говорит о том, что основной кластер, узлы которого обладают большими значениями центральности, очень неоднороден и обладает очень сложной структурой в связи с низким коэффициентом кластеризации. Также можно сделать вывод, что реальные структуры, характеризующиеся наибольшей алгебраической

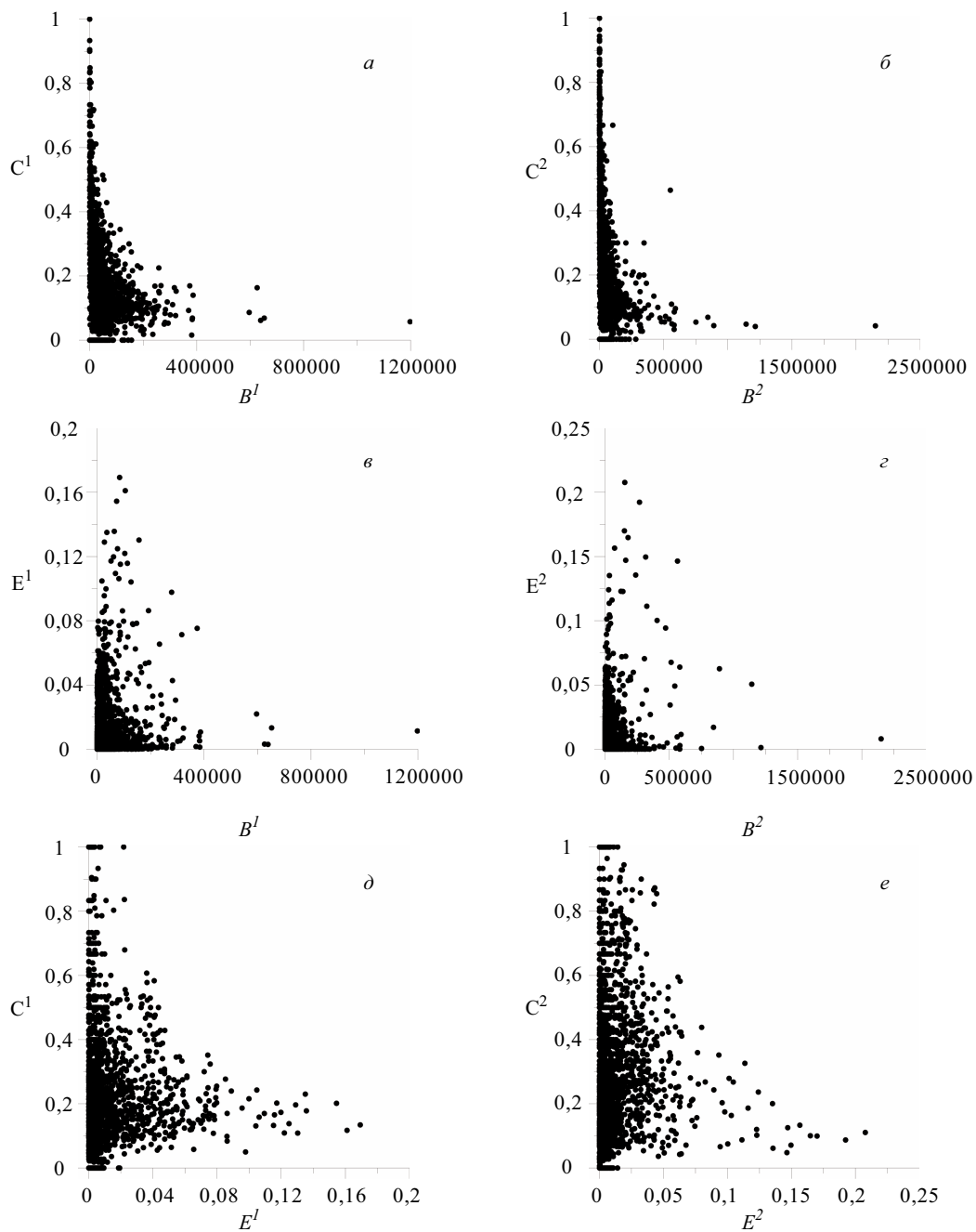


Рис. 3: Зависимости рассчитанных параметров друг от друга: *a, v, d* — для сети, построенной на принципе жанровой принадлежности, *b, z, e* — сети коллабораций

связностью, возникают вследствие сложного взаимодействия между узлами сети, и порождают сильно неоднородные структуры в её топологии.

### ЗАКЛЮЧЕНИЕ

В данной работе были изучены процессы кластеризации в двух социальных сетях — коллабораций и жанровой принадлежности музыкантов — облада-

ющих идентичными наборами узлов, но разной топологией связей. В результате были выявлены схожие показатели рассчитанных параметров для структур одной природы, что позволяет говорить об определенной универсальности полученных зависимостей и возможности их применения для исследования процессов формирования топологий в дальнейших исследованиях.

Работа выполнена при финансовой поддержке РФФИ (проект 16-32-00334).

- 
- [1] *Onnela J.P., Saramäki J., Hyvönen J., Szabo G., Lazer D., Kaski K., Kertész J., Barabasi A.-L.* Proc. Nat. Acad. Sci. USA. 2007. **104**. P. 7332.
- [2] *Valencia M., Martinerie J., Dupont S., Chavez M.* Phys. Rev. E. 2008. **77**, N5. P. 050905.
- [3] *Ulhaas P.J.* Frontiers Neurosc. 2009. **3**, P. 17.
- [4] *Huang R., Sun X.* Information, Communication & Society. 2014. **17**, N1. P. 86.
- [5] *Estrada E., Meloni S., Sheerin M., Moreno Y.* Phys. Rev. E. 2016. **94**, N5. P. 052316.
- [6] *Barabasi A.L., Oltvai Z.N.* Nature reviews genetics. 2004. **5**, N2. P. 101.
- [7] *Shakkottai S., Srikant R.* Foundations and Trends® in Networking. 2008. **2**, N3. P. 271.
- [8] *Stehlé J., Barrat A., Bianconi G.* Phys. Rev. E. 2010. **81**, N3. P. 035101.
- [9] *Wang Z., Jusup M., Wang R.W., Lei S., Isawa Y., Moreno Y., Kurths J.* Science Advances. 2017. **3**, N3. P. e1601444.
- [10] *Cuesta J.A., Gracia-Lázaro C., Ferrer A., Moreno Y., Sanchez A.* Scientific reports. 2015. **5**. P. 7843.
- [11] *Robins G., Snijders T., Wang P., Handcock M., Pattison P.* Social networks. 2007. **29**, N2. P. 192.
- [12] *Scott J.* Sociology. 1988. **22**, P. 109.
- [13] <http://www.allmusic.com>.
- [14] *Callaway D.S., Newman M.E., Strogatz S.H., Watts D.J.* Phys. Rev. Lett. 2000. **85**, N25. P. 5468.
- [15] *Barabási A.L., Albert R., Jeong H.* Physica A: Statistical mechanics and its applications. 2000. **281**, N1. P. 69.
- [16] *Wasserman S., Faust K.* Cambridge university press. 1994. **8**.
- [17] *Bonacich P.* J. of Math. Soc. 1972. **2**, N1. P. 113.
- 

## Formation of clusters in a complex network reflecting the process of cooperation in social groups

V. O. Nedajvozov<sup>a</sup>, V. V. Makarov<sup>b</sup>, M. V. Goremyko<sup>c</sup>, D. V. Kirsanov<sup>d</sup>

<sup>1</sup>Yuri Gagarin State Technical University of Saratov. Saratov 410056, Russia  
E-mail: <sup>a</sup>fdf\_het@mail.ru, <sup>b</sup>vladmak404@gmail.com, <sup>c</sup>gormv67@mail.ru, <sup>d</sup>dankirsdot@gmail.com

In this study, we investigated the processes of clustering in multilayered social networks of collaborations and professional interactions. We have identified the basic principles that are responsible for the interconnection of topology and dynamic characteristics in such kind of networks.

PACS: 05.45.-a

*Keywords:* complex network theory, social network, oscillator, nonlinear oscillations.

*Received 23 July 2017.*

### Сведения об авторах

- Недайвозов Владимир Олегович — студент; e-mail: fdf\_het@mail.ru.
  - Макаров Владимир Владимирович — канд. физ.-мат. наук, ассистент; e-mail: vladmak404@gmail.com.
  - Горемыко Михаил Владимирович — руководитель управления имуществом комплексом; e-mail: gormv67@mail.ru.
  - Кирсанов Даниил Владимирович — студент; e-mail: dankirsdot@gmail.com.
-