

РАЗРАБОТКА КРИТЕРИЕВ БЕЗОПАСНОСТИ СЛОЖНЫХ СИСТЕМ ЭЛЕКТРОННОГО ДЕНЕЖНОГО ОБОРОТА С ПОМОЩЬЮ ГРАФОВОЙ АНАЛИТИКИ

Пыркина О.Е., Зададаев С.А.

Финансовый университет при Правительстве РФ
Россия, 125993 (ГСП-3), г. Москва, Ленинградский просп., 49

OPyrkina@fa.ru, SZadadaev@fa.ru

Аннотация: представлена графовая модель для исследования системы электронного денежного оборота посредством применения теории сложных систем. На основе моделирования разрушения системных связей построена методика оценки устойчивости системы, позволяющая по метрикам графа сделать вывод о степени разрушения системы и эффективности ее функционирования.

Ключевые слова: электронный денежный оборот, сложные системы, графовая аналитика.

Введение

Теория сложных сетей широко применяется в последние годы в самых разных научных областях, в том числе в экономике и финансах [1]. Так, графовый подход к математическому моделированию успешно применяется для исследования неопределенности финансового рынка, в теории корпоративного управления, в инвестиционном анализе [2], но ранее такие методы не применялись для исследования и мониторинга систем электронного денежного оборота. Пилотное исследование авторов в этой области [3], опубликованное в 2019 году, положило начало целому ряду численных экспериментов по графовому моделированию систем электронного денежного оборота.

Применение парадигмы сложных сетей вполне оправдано во многих сферах в связи со все возрастающей цифровизацией всех процессов современной жизни в целом и финансовой деятельности в частности. Такая цифровизация требует исследования все большего числа имеющихся одновременно связей между объектами (экономическими агентами). Одну из наиболее перспективных возможностей такого анализа и предоставляет теория сложных сетей, в частности, является графовая аналитика.

В современной банковской и финансовой практике используется большое количество различных систем для осуществления электронного денежного оборота, включая системы электронной коммуникации между банком и его клиентами, банков между собой, а также коммуникации между финансовыми учреждениями различных стран. Такого рода системы обладают всеми свойствами технологически сложных систем, включая наличие концентраторов - хабов. Естественно отображать такие системы с помощью графа и исследовать их свойства и характеристики методами графовой аналитики

1 Основные характеристики модели

1.1 Общее описание модели

Система электронного денежного оборота отображается графом, вершинами которого изображают участников электронного денежного оборота, а ребра – имеющие между участниками связи и возможные транзакции. Для простоты все связи между участниками отображаются одним ребром. Участники электронного денежного оборота классифицируются как компании (B), государственные организации (G), банки и межбанковские центры обработки платежей (B), таким образом в модели исследуются связи типов B2B, B2G, G2B, G2G. Электронные переводы денежных средств между частными лицами (P2P) и соответствующие им связи в модель не включены, потому что авторы полагают, что их возможное нарушение (разрушение) не приводит к существенному нарушению функционирования системы в целом

Начальная модель была разработана на основе выборок, включающих в себя не более 300 вершин. Авторы опирались в этом выборе на исследования Сбербанка России по построению графовых моделей для кредитного скоринга, представленные на научно-практической конференции «Scoring days 2018» С. Голицыным. В этих исследованиях было эмпирически показано, что при увеличении числа вершин свыше 300 граф становится насыщенным, и его метрики далее не меняются при увеличении числа включенных в него вершин. (<https://drive.google.com/drive/folders/1RAz3FQx96qkenEnd7ne3GFnlS4mThsgu>, дата обращения 12.05.2018). Имитационное моделирование с большим числом вершин, проведенное авторами, также не показало изменения характеристик сети при увеличении числа вершин графа.

Графовая модель строилась на основе набора вершин и ребер (такой набор содержит необходимую информацию обо всех участниках денежного оборота). В этот набор включались следующие атрибуты:

- Для государств - коды по стандарту ISO 3166
- Для банков - коды SWIFT (в международных клиринговых расчетах), информация о владельцах, номера счетов и т.д.
- Для юридических лиц – идентификаторы ИНН, данные о владельцах, номера счетов, категория компании и т.д.

Для создания графа по этим атрибутам использовался свободно распространяемый программный пакет Gephi, дающий возможность найти общие атрибуты участников электронного денежного оборота и при наличии таких общностей создать дуги, соединяющие вершины с общими атрибутами, то есть подтвердить их связь. Кроме того, использование пакета позволяло вычислить числовые значения различных метрик графа и визуализировать нужные фрагменты графа.

На Рис. 1 представлен один из типовых графов, сгенерированный программой Gephi. Этот граф сгенерирован для сложной сети, состоящей из 300 экономических агентов, соединяемых 2197 ребрами (дугами), что отражает финансовые взаимоотношения и транзакции между ними (включая коммуникации вида G2B/B2G и B2B)

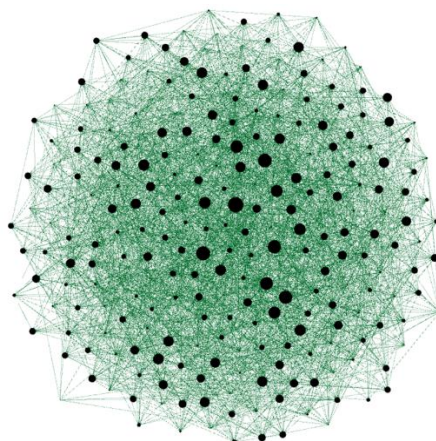


Рис. 1. Пример типичного графа. Сгенерированного программой Gephi

1.2 Основные метрики графа, используемые при моделировании

К метрикам центральности графа традиционно относят метрики, агрегирующие информацию о взаимном расположении его вершин и ребер. Понятие центральности определяет степень важности конкретной вершины графа и ее «влияния» на окружающие структуры графа. Математические методы для подобных метрик центральности активно разрабатывались в последние десятилетия для анализа социальных сетей (Social Network Analysis, SNA). Числовые значения этих метрик дают много информации для идентификации сообществ и активно используются на практике, например, для обнаружения возможной террористической активности в сети интернет [4, 5].

Наибольшую эффективность для оценки устойчивости сети показали при моделировании следующие метрики центральности графа:

1. *Центральность по степени*, или степень центральности (*Degree centrality*), которая рассчитывается на основе матрицы связности $[a_{ij}]$ графа G . Вершины с высоким значением этого показателя имеют существенную активность и информационное влияние на ближайших соседей. Метрика вычисляется как

$$D_i = \frac{k_i}{N-1} = \frac{\sum_{j \in G} a_{ij}}{N-1} \quad (1)$$

и равна количеству смежных вершин k_j , нормированному на максимально возможное их количество $N-1$.

2. *Центральность по посредничеству*, или степень промежуточности (*Betweenness centrality*), рассчитывается как

$$B_i = \frac{\sum_{i < j \in G} n_{jk}(i) / n_{jk}}{(N-1)(N-2)} \quad (2)$$

где n_{ij} есть количество всех кратчайших путей между узлами j и k , $n_{ij}(i)$ - количество таких кратчайших путей между узлами j и k , проходящих через узел i . Вершины графа с высоким значением этой метрики осуществляют обмен информацией между несколькими подмножествами графа, разрушение таких вершин может нарушить функционирование всей сети.

3. *Центральность по близости*, или степень близости (closeness centrality), эта метрика вычисляется как

$$C_i = (L_i)^{-1} = \frac{N-1}{\sum_{j \in G} d_{ij}} \quad (3)$$

где d_{ij} есть дистанция между узлами i и j , и L_i - нормированное расстояние между узлом i и другими вершинами графа. Если эта метрика имеет низкое значение, путь через соответствующий узел позволяет быстро достичь нужных вершин графа.

4. *Центральность собственного вектора*, или метрика собственного вектора (eigenvector centrality), вычисляется как

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j \in G} a_{ij} x_j \quad (4)$$

где $M(i)$ обозначает набор соседей вершины i и λ - собственное значение вектора

5. *Метрика*, или *ранг*, *индекса* (PageRank) позволяет провести сравнение относительной важности вершин графа, вычисляется рекуррентным образом как

$$PR(A) = (1-d) + d \sum_{B \in M(A)} \frac{PR(B)}{L(B)} \quad (5)$$

где $M(A)$ - множество соседей узла A , $L(B)$ - количество связей, исходящих из узла B , $d \in (0,1)$ - демпфирующий фактор (константа).

1.3 Ранжирование и построение интегральной метрики центральности для оценки устойчивости сети электронного денежного оборота

На основе перечисленных и описанных выше графических метрик (формулы (1)-(5)) была разработана интегральная метрика для оценки устойчивости сети. Она основана на построении ранговых распределений возможных величин метрик, и ее числовые значения позволяют сделать вывод о степени устойчивости сети

В модели применялась следующая процедура построения ранговых распределений:

1) диапазон значений каждой перечисленной и описанной выше метрики графа, рассчитанной программой Gephi, разбивается на четыре квартиля стандартным образом по правилам математической статистики;

2) для метрик центральности D_i , B_i , x_i и $PR(i)$:

а) вершинам с номером i графа, значения метрик которых попадают в верхний квартиль, присваивается наивысший уровень важности, их ранг равен 3;

б) вершинам с номером i графа, попадающим в межквартильный диапазон присваивается средний уровень важности, их ранг равен 2;

в) вершинам, попавшим в первый квартиль, присваивается низший уровень важности, их ранг равен 1.

3) для метрики центральности C_i ранговое распределение строится обратным образом:

а) наивысший уровень важности с рангом 3 присваивается вершинам, значение метрики которых попадет в нижний квартиль,

б) средний уровень с рангом 2 присваивается вершинам со значением метрики в межквартильном диапазоне;

в) низший уровень важности с рангом 1 присваивается узлам со значением метрики в верхнем квартиле.

4) На основе описанного в пунктах 2 и 3 алгоритма построения ранговых оценок (ранжирования) строится итоговая количественная метрика ранга важности узла (вершины) с номером по пяти метрикам центральности графа как

$$rang(i) = rang_{D_i}(i) + rang_{B_i}(i) + rang_{C_i}(i) + rang_{x_i}(i) + rang_{PR(i)}(i) \quad (6)$$

Такие вычисления проводятся для всех вершин графа.

5) Проводится сортировка по убыванию полученных рангов всех узлов (вершин) графа, диапазоном изменения рангов является промежуток от 5 до 15; это позволяет классифицировать вершины графа по уровню важности для бесперебойного функционирования системы оборота электронных денег.

а) Вершины в верхней части ранжированного перечня важности с высокими значениями метрики ранга важности 15, 14 и 13 оказывают наибольшее влияние на устойчивость функционирования всей системы, сбой в этих узлах с высокой вероятностью приведут к нарушению функционирования платежной системы в целом; они определяют «красную зону»

б) Вершины в средней части ранжированного перечня важности с промежуточными значениями метрики ранга 12, 11, 10, 9, 8 попадают по степени влияния на безопасность в «желтую зону»

в) Вершины с низшими значениями метрики ранга 7, 6 и 5 обладают минимальным влиянием на безопасность функционирования системы, их относим к «зеленой зоне»

б) Если система моделируется единым графом, то дальнейший вывод об оценке устойчивости сети строится с помощью следующей Матрицы устойчивости (Таблица 1)

Таблица 1. Матрица устойчивого функционирования и стабильности сети

Доля вершин графа в «желтой» зоне	Доля вершин графа в «красной» зоне		
	< 25%	25-50%	>50%
<50%	Благоприятный уровень	Пограничный уровень	Неприемлемый уровень
>50%	Пограничный уровень	Неприемлемый уровень	

7) Если же система моделируется несколькими графами для различных своих фрагментов, тогда для каждого фрагмента подсчитывается число вершин в «красной», «желтой» и «зеленой» зонах. Далее строится усредненная оценка процентных соотношений числа вершин в каждой зоне с весовыми коэффициентами, определяемыми долей количества узлов каждого фрагмента по отношению ко всей системе в целом. Далее вывод о приемлемости уровня устойчивости делается на основе той же матрицы (Таблица 1)

2 Оценка эффективности модели с помощью краш-тестов

Оценка эффективности модели проводилась с помощью разработанных авторами краш-тестов. Эти краш-тесты включали в себя моделирование последовательного разрушения вершин графа, выбранных в случайном порядке, и отслеживания качества прохождения транзакций в такой «нарушенной» сети электронного денежного оборота. Разрешение сети, результатом которого является исключение экономических агентов из оборота электронных денежных средств, моделирует возможные технические сбои и хакерские атаки.

Схема моделирования включает генерирование случайного графа для сложной сети с различными характеристиками вершин (включая концентраторы-хабы) и различными ребрами, отображающими финансовые взаимодействия между ними. При последовательном, шаг за шагом, удалении вершин графа отслеживалась эффективность прохождения и завершения транзакций. Наиболее убедительное графическое представление сгенерированного графа и зависимость доли завершённых транзакций от доли разрушенных вершин представлены на Рис 2 и Рис.3.

Рис.2 представляет для иллюстрации пример графа со следующими параметрами: не более 60 консорциумов банков (хабов). Включающих в себя не более 100 банков, и не более 1000 корпоративных клиентов банков (экономических агентов, участников электронного денежного оборота)

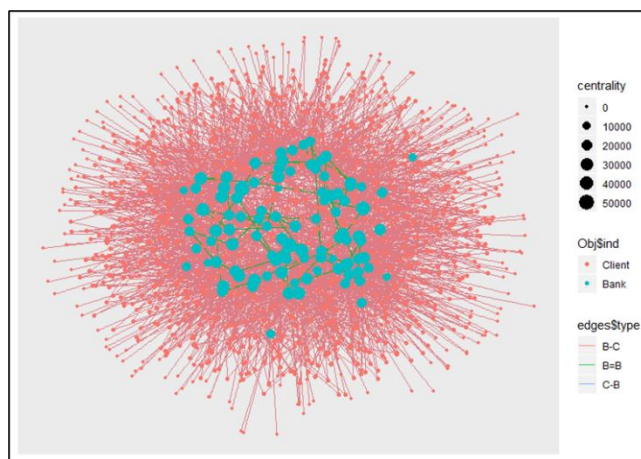


Рис. 2. Граф сложной сети электронного денежного оборота со случайными вершинами и ребрами, используемый в краш-тестах

Вероятности того, что экономический агент имеет счета в одном, двух, трех или четырех банках задавались вектором (0.2, 0.4, 0.3, 0.1). Векторы с другими значениями вероятностей также рассматривались при моделировании, но они не изменили результатов в существенной степени.

Контроль функционирования сложной сети электронного денежного оборота осуществлялся путем вычисления доли успешно завершенных транзакций для всех типов вершин. Рассматривались транзакции, завешенные за 2, 3, 4 или 5 шагов.

Результаты моделирования для графа, представленного на Рис 2, изображены на Рис.3. Синяя линия задает общую осуществимость транзакции. (за 2, 3, 4 или 5 шагов), то есть долю транзакций, которые успешно завершены. Красная линия представляет долю вершин в «красной» зоне в процессе осуществления транзакции, желтая линия, соответственно, долю вершин в «желтой» зоне, вертикальная черная линия определяет границу области осуществимости транзакции. Результаты моделирования подтверждают критерии, приведенные в матрице устойчивости системы (Таблица 1), и позволяют сделать вывод о корректности модели

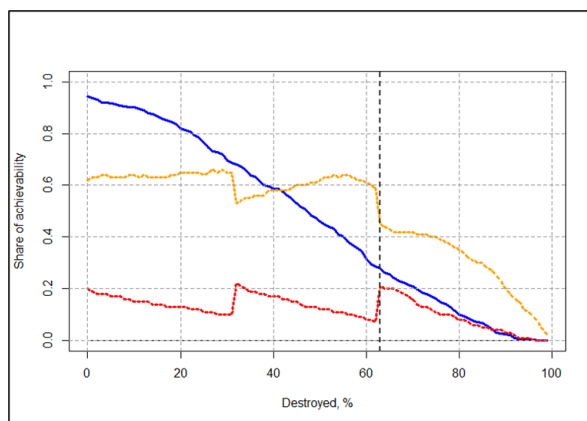


Рис. 3. Зависимость осуществимости транзакции для графа, представленного на Рис.2, и доля вершин в различных зонах устойчивости в зависимости от доли разрушенных вершин.

Следует отметить, что схема последовательного разрушения вершин осуществлялась ранее в совершенно иных задачах математического моделирования [6], таких как распространение кризисных явлений в экономике, а не в исследованиях стабильности финансовых систем.

Аналогичные результаты были получены и для графов иной структуры, пример такого графа приведен на Рис. 4. и Рис.5.

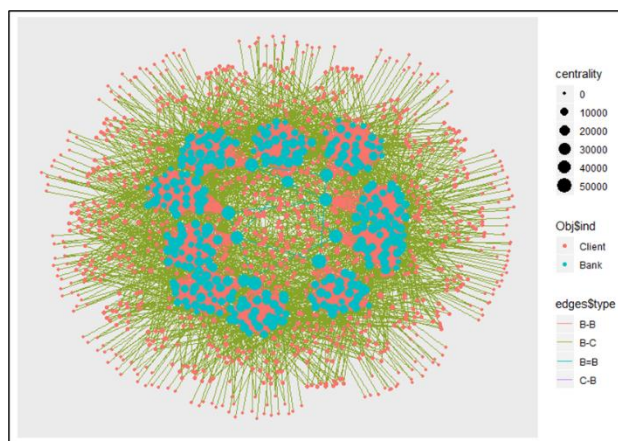


Рис. 4. Граф электронного денежного оборота для системы со следующими параметрами: не более 10 консорциумов банков (хабов), включая не более 300 банков и не более 1000 корпоративных клиентов банков (экономических агентов). Вероятности, что экономический агент имеет счета в одном, двух или трех банках описываются вектором (0.2, 0.4, 0.3, 0.1).

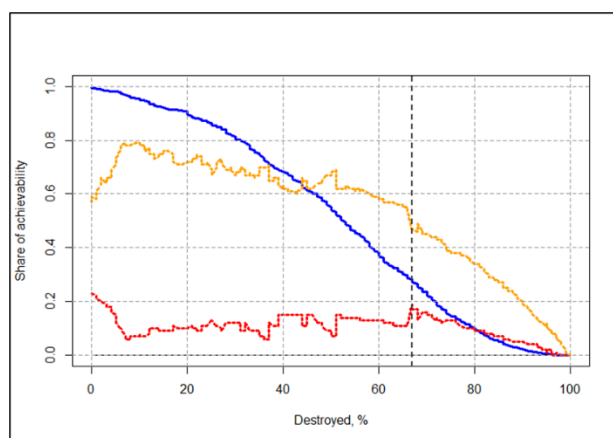


Рис. 5. Зависимость осуществимости транзакции для графа, представленного на Рис.4, и доля вершин в различных зонах устойчивости в зависимости от доли разрушенных вершин.

Заключение

В работе представлена разработанная авторами методика мониторинга и анализа бесперебойного функционирования электронного денежного оборота. Система такого денежного оборота рассматривается как сложная сеть, моделируемая графом и исследуемая методами графовой аналитики. Наибольшую эффективность при моделировании показали метрики центральности графа, на основе которых и была построена интегральная метрика.

Краш-тесты, разработанные авторами, позволили провести валидацию модели и предложенной методики, а также найти критерии бесперебойного функционирования сети электронного денежного оборота. Критерии основаны на изменении в метриках центральности по мере увеличения доли разрушенных вершин графа. Такое разрушение успешно имитирует реальные процессы, приводящие к ограничению прохождения информации по сети (прохождению транзакции). Идентифицирован критический порог разрушения, по достижении которого система прекращает устойчивое функционирование.

Методика, разработанная авторами на модельных структурах, может быть применена к графам, основанным на реальных данных, для мониторинга и контроля устойчивости и бесперебойного функционирования сети электронного денежного оборота

Литература

1. *V. A. Kalyagin, P. Pardalos, Th. M. Rassias*, Network Models in Economics and Finance, Springer International Publishing, 2014
2. *A. M. D'Arcangelis, G. Rotundo*, Complex Networks in Finance. Complex Networks and Dynamics, Lecture Notes in Economics and Mathematical Systems, Springer International Publishing Switzerland, 2016
3. *E. Pyrkina and S. A. Zadadaev*, "Application of Crash-Tests for Graph Analytics of Electronic Payment Complex System for Investigation of System Economic Security," in 2019 Twelfth International Conference "Management of large-scale system development" (MLSD), Moscow, Russia, 2019
4. *P. Choudhary*. A Survey on Social Network Analysis for Counter-Terrorism. International Journal of Computer Applications 112(9), 2015, DOI: 10.5120/19695-1455
5. *Basarab M.A., Ivanov B.G., Kolesnikov A.V, Matveev V.A.* Detection of illegal activities in cyberspace on the basis of social network analysis. Algorithms, Methods and Tools. (a survey). Voprosy cyberbezopastnosti (in Russian, with abstracts in English) №4(17) - 2016
6. *Caetanoa M.A.L, Yoneyama T.* The effects of node exclusion on the centrality measures in graph models of interacting economic agents. Physica A 430 (2015) 216–223.