

САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

Дисциплина: Анализ социальных сетей

Отчёт  
по лабораторной работе  
АНАЛИЗ СЕТИ СООБЩЕСТВА

Выполнили: Анастасия Козлова, Мария Ларионова,  
Анна Ревенко, Марина Бережнова  
К3443

Проверила Н. Ф. Гусарова

2021 г.

Задача: выбрать сообщество; собрать информацию о сообществе; проанализировать активных пользователей; вывести график активности реакций; построить тематический профиль постов сообщества; проанализировать сеть общения.

Выполнение:

Для анализа выбрано сообщество фанатов молодой петербургской группы Super Collection Orchestra.

Сообщество интересно для анализа тем, что коэффициент связанности между его участниками очень высокий, и это происходит не за счёт наличия звёзд-лидеров, а за счёт существования сети общения с многочисленными узлами. Также анализ данного сообщества будет основываться на событийной модели, т.е. исследование может быть построено на мониторинге математической модели реакций на реальные изменения в группе.

Узнаем первичную информацию о группе:

Количество подписчиков	4299
Количество публикаций	428
Количество активных пользователей	198
Доля активных пользователей	~5%

Domain = supcolorc

Owner\_id = 124540357

*Вывод активных пользователей:*

```
library(vkR)
library(tm)
library(stringr)
library(lda)
library(LDAvis)
library(servr)
vkoAuth(7643106)
setAccessToken("a29a977849e2955112e74473e4eaa62a8542f42dba369ff0baa0045ad5eb2572d56f4194c681b89703e97")

members <- getGroupsMembersExecute('supcolorc')
users <- getUsersExecute(members)

wall <- getWallExecute(domain = "supcolorc", progress_bar = TRUE)
comments <- wallGetCommentsList(wall, progress_bar = TRUE)
users_commented <- sapply(comments,
                           function(comment) data.frame(from_id = comment$from_id))
users_commented <- do.call(rbind, users_commented)
likes <- sapply(wall$posts$id,
                function(post_id) likesGetList(type = 'post', owner_id = '-124540357', item_id = post_id))
users_liked <- getUsersExecute(unlist(likes["items",]))
users_posted <- data.frame(from_id = wall$posts$from_id)
active_users <- getUsersExecute(unique(c(users_commented$from_id, users_posted$from_id, users_liked$id)))
nrow(active_users)
```

Любопытно отметить, что бывшие участники группы больше не являются активными пользователями паблика. В 2019 году группу покинул Дмитрий Батарин, чуть позже – Андрэ Манолли, но, несмотря на их сравнительно недавнюю сепарацию, они больше не проявляют активного интереса в жизни Super Collection Orchestra.

## Активные пользователи сообщества (музыканты выделены красным цветом):

	first_name	id	last_name	25	Лёша	125557513	Кож	50	Ахе	40087	Relaxe	75	Иван	3508412	Саченко
1	Владислав	124626793	Шадрин	26	Маша	167428450	Брумберг	51	Ekaterina	163344	Golubeva	76	Александр	3702810	Коновалов
2	Хорошо	5432615	Очень	27	Филипп	70372787	Брусковский	52	Тоня	182160	Михайлова	77	Роман	3774840	Шагалиев
3	Ксения	203921572	Савицкая	28	Диана	10356772	Бестужева	53	Александр	235839	Ющенко	78	Дарья	3834179	Макарова
4	Эмилий	19784900	Плясов	29	Оля	24604264	Плохотнюк	54	Денис	701371	Харлашин	79	Виктор	4420106	Филатов
5	Ярослава	12246095	Миненкова	30	Настя	35766941	Мазилова	55	Иван	850814	Зализко	80	Татьяна	4538095	Борисенко
6	Alina	99692554	Balk	31	Виктория	70730800	Васильева	56	Денис	851468	Рубин	81	Doos	4881283	Doos
7	Жанна	232550424	Лыжина	32	Tolik	504792	Voronetskii	57	Роман	929851	Виноградов	82	Иван	5265034	Шуваев
8	Марк	396844024	Мерекко	33	Сергей	2376935	Кудрящеев	58	Илья	1135930	Egorov	83	Большая	5915025	Матуся
9	Герман	206904150	Коваленков	34	Юра	370012333	Аристов	59	Shooting	1390664	Star	84	Юлия	6343717	Ишиева
10	Jims	1160977	Ahagov	35	Анастасия	288359613	Бибикина	60	Сергей	1406300	Винник	85	Олег	6504627	Шарков
11	Ольга	14028282	Забаява	36	Александр	5670639	Фламин	61	Виктор	1467717	Ужаков	86	Дарья	6893878	Титова
12	Денис	210315595	Белых	37	Александр	371331	Струкалёв	62	Даниил	1511627	Романович	87	Сурен	7274143	Саркисян
13	Дмитрий	1255882	Паутов	38	Антон	50675895	Гурский	63	Vaiya	1686457	Volkova	88	Сергей	7335123	Снегирев
14	Георгий	254869350	Жуков	39	Ян	343844175	Каменский	64	Степан	1740036	Бызов	89	Андрей	8060013	Пучков
15	Юра	4788345	Алемовский	40	Ксюшок	4172539	Меньщикова	65	Дмитрий	1744684	Юшин	90	Владислав	8436280	Гончаров
16	Денис	380485661	Лебедев	41	Фёдор	342081572	Денисов	66	Даниил	2182297	Непёный	91	Vadim	8591339	Took
17	Игорь	90329256	Гинзбург	42	Андрей	421279344	Григорьев	67	Julia	2240476	Fruktenbein	92	Кирилл	8604165	Коперник
18	Александр	21669724	Рогов	43	Алексей	55252348	Плотников	68	Евгений	2302285	Роднянский	93	Наталья	9000586	Бугенко
19	Саша	9027114	Строганов	44	Василиса	4251296	Сталь	69	Рома	2311900	Реднек	94	Сергей	9292494	Каверин
20	Антон	1543633	Максимов	45	Никит	4770087	Дмитриев	70	Женя	2565868	Кит	95	Иван	9510911	Лихачёв
21	Елена	50142100	Логинава	46	Андрей	17392929	Гриценко	71	Илья	2682374	Sergееv	96	Ильдар	11600456	Касимов
22	Сергей	324473	Седаков	47	Игнат	252887	Иконников	72	Евгений	2690623	Барташевич	97	Анастасия	11844065	Воронина
23	Лиза	27751544	Новых	48	Raul	22155	Mikhailov	73	Анастасия	3114153	Смирнова	98	Алена	12319701	Сорокина
24	Екатерина	265104875	Шикова	49	Наталья	35547	Ашуровская	74	Живодкин	3474721	Макс	99	Яков	12427438	Свиридов

100	Ира	12789962	Садкова	125	Салсалыч	80732985	Поздюша	150	Алёна	80587004	Зверева	175	Artur	235523357	Flint
101	Катерина	13869119	Захарова	126	Кристина	46413519	Журавлёва	151	Александра	83644440	Шараева	176	Андрей	244828749	Михайлов
102	Олег	14063022	Старожук	127	Dania	241532905	Konoplov	152	Евгений	88204911	Есиков	177	Ася	457968753	Бочкарева
103	Сергей	14237324	Герасимов	128	Тимофей	20507920	Коляда	153	Роман	108311892	Жирнов	178	Horse	8688557	Complex
104	Наталья	14687876	Сударикова	129	Илья	41952664	Иванов	154	Лилия	120529743	Мансурова	179	La	79159129	Rupture
105	Данила	15889538	Данилов	130	Полина	52842574	Чудинова	155	Полина	149723856	Королёва	180	Евангелина	367444147	Черноусова
106	Polina	17757422	Afanasieva	131	Андрей	68066243	Квачёв	156	Светлана	169113036	Завьялова	181	Никита	247306425	Камышев
107	Елена	18251034	Паутова	132	Катя	185296300	Коноплева	157	Юля	184257975	Разумовская	182	Никита	881881	Акулов
108	Элина	19443356	Хажеева	133	Ирена	296381774	Нарбутовских	158	Sayam	199805601	Sounds	183	Дарья	19625899	Чупилина
109	Наталья	20078463	Рыбакина	134	Ник	300688805	Брусковский	159	Станислав	259773233	Козалев	184	Fedor	32572540	Samoilov
110	Elijah	20119800	Deadwood	135	Сергей	360490481	Ломакин	160	Наталья	349996723	Белая	185	Людмила	62810367	Кендалиева
111	Uyt	21810209	Ruin	136	Евгений	774583	Овцебокс	161	Liesbeth	430740645	Salander	186	Майя	189272221	Кава
112	Роман	23105470	Васильев	137	Виктория	1072940	Панфилова	162	Egor	483722657	Мельник	187	Peels	471542698	Repeek
113	Сашка	24972142	Новиков	138	Дарья	1073348	Дарья	163	Лиза	486630192	Соколова	188	Ioane	483915105	Asabashvili
114	Lera	25753232	Purtova	139	Екатерина	1642918	Клементьева	164	Наталья	528637421	Бугайченко	189	Мила	586595131	Мижеева
115	Николай	28613689	Ордановский	140	Павел	3737559	Евгеньевич	165	Степан	1189589	Казарьян	190	Федор	29406407	Вылегжанин
116	Борис	28792761	Штольц	141	Наталья	7090182	Гусенкова	166	Святослав	13176782	Щедромирский	191	Таня	31806001	Коленова
117	Алиса	204848807	Стрельцова	142	Светлана	10484687	Есина	167	Алексей	28837359	Кусков	192	Ксения	394778	Седнина
118	Никита	111375400	Осиев	143	Ульяна	13405035	Тряпичкина	168	Алексей	59475528	Мочалов	193	Sonia	3636889	Kovitz
119	Настя	175624557	Мамолкина	144	Мария	18477383	Жураковская	169	Элиза	95035916	Филиппова	194	Илья	34049640	Малыгин
120	Михаил	82133452	Комаров	145	Nikita	20447623	Ivaschenko	170	Гульназа	112379223	Халикова	195	Алекс	61264383	Рубиков
121	Valkosipuli	515625042	Tomaattimurska	146	Дарья	36855335	Щёлокова	171	Никита	151620828	Макаров	196	Анастасия	66275932	Сергеева
122	Динеш	173945	Лекамге	147	Анастасия	37171271	Цветкова	172	Фрэнк	167419784	Айеро	197	Саша	258566314	Лавищенко
123	Екатерина	136525041	Катышева	148	Serge	37570292	Gam	173	Варвара	196383315	Овчарова	198	Polina	419222347	Belyavskaya
124	Юрий	4353499	Коваленко	149	Александр	73785643	Хромов	174	Алексей	217560941	Вахрамеев				

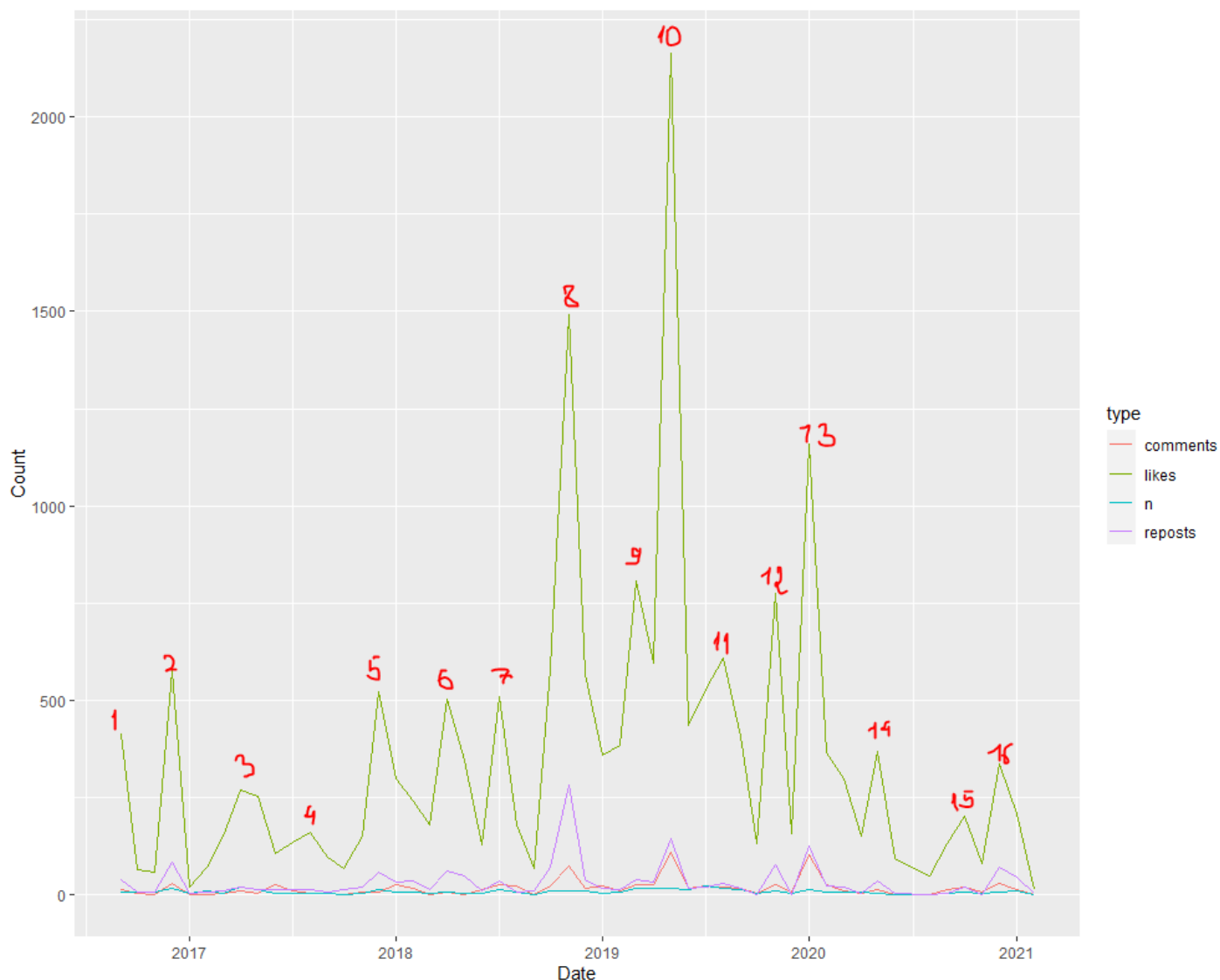
Построим график активности подписчиков группы.

```
domain <- 'supcolorc'
wall <- getWallExecute(domain = domain, count = 0, progress_bar = TRUE)
metrics <- jsonlite::flatten(wall$posts[c("date", "likes", "comments", "reposts")])
metrics$date <- as.POSIXct(metrics$date, origin="1970-01-01", tz='Europe/Moscow')

library(dplyr)
df <- metrics %>%
  mutate(period = as.Date(cut(date, breaks='month')))) %>%
  group_by(period) %>%
  summarise(likes = sum(likes.count), comments = sum(comments.count), reposts = sum(reposts.count), n = n())

install.packages("ggplot2")
install.packages("tidyr")
library(ggplot2)
library(tidyr)
ggplot(data=gather(df, 'type', 'count', 2:5), aes(period, count)) + geom_line(aes(colour=type)) +
  labs(x='Date', y='count')
```

График активности реакций (всплески активности пронумерованы):



Можно попытаться выделить предполагаемые причины всплесков активности:

- 1) Релиз песни
- 2) Релиз песни
- 3) Лайв с выступления и релиз песни
- 4) Интервью
- 5) Плейлисты от музыкантов
- 6) Релиз песни
- 7) Плейлисты от музыкантов
- 8) Официальный релиз альбома
- 9) Видео с презентации нового альбома
- 10) Официальный релиз альбома
- 11) Участие в фестивале «Боль»

- 12) Официальный релиз сингла
- 13) Клип
- 14) Официальный релиз альбома
- 15) Официальный релиз песни к трейлеру фильма
- 16) Лайв с фестиваля

С начала 2020 мы наблюдаем спад активности реакций – можно предположить, что причиной тому послужила пандемия, во время которой, как известно, любые выступления были отменены. При анализе роста популярности группы до пандемии возможен вывод о том, что после снятия ограничений возрастающая тенденция снова начнёт преобладать.

Построим тематический профиль постов Super Collection Orchestra с использованием алгоритма k-means.

*Загрузка всех постов группы и очистка получившегося массива текста от ненужных символов*

```
posts <- c()
posts <- c(wall, as.vector(posts$body))

# Очищаем текст от "мусора"
posts <- str_replace_all(posts, "[ē]", "e")
posts <- str_replace_all(posts, "[[:punct:]]", " ")
posts <- str_replace_all(posts, "[[:digit:]]", " ")
posts <- str_replace_all(posts, "http\\s+\\s*", " ")
posts <- str_replace_all(posts, "[a-zA-Z]", " ")
posts <- str_replace_all(posts, "\\s+", " ")
posts <- tolower(posts)
posts <- str_trim(posts, side = "both")

write(posts, file='posts_processed.txt')
```

*Фрагмент получившегося массива:*

— = = = = премьера мы решили что лучший способ представить новый сингл сразу снять на него клип клип создан в рамках проекта глобальной инициативы бренда | в поддержку артистов художников барме а ника про музыку о под эту музыку можно даже послать но | января нам точно не удастся поспать нас ожидает новый день без сна который мы поведем в новом клубе время билеты на концерт тут пока вы до вас особо не беспокоить и не подогревать хотя было чем мы трудились и очень скоро покажем вам новый клип пока все спасибо за внимание < + > новости съездили в москву переобуться в заодно нахлобучил кадр | булочки поиск + > мала мальчика в моменте выбора нового сценического костюма киномены здесь совместно с | мы написали песню к трейлеру фильма | майор гром чумной доктор но мы не придумали назван т с новыми песнями < + > месток + > тонн арбат стартс + > билеты можно приобрести тут < + > так бывает нам нужна ваша поддержка все что накопили на югах за время паломничества отдадим воскресным ве с и так появилась группа | ребята уже активно пишут музыку снимают клипы и выступают мы рады за них и следим за их творчеством < + > наш состав пополнился новыми музыкантами и сегодня мы хотим под одаривает ослепительной улыбкой воды смыкаются вновь это филоп дамы и господа все смешалось со всем петербург с берлином лето с зимой женщины мужчины кони потом как то снимали лайв а ник гнал нас ьсь и волка ассистент режиссера | ольга забава всех организовала накормила согрела принесла лопаты | иван зализко | станислав рассолов павел азаров евгений прокопович строили завешивали дымили светили ретий пришло время прочтаться со страхами представляем последний мини фильм из серии < + > режиссер | в ролях | наташа чернявская | даша чернявская при поддержке ленфильма не бойсь фильм второй < + > в прямые эфиры с нашими друзьями и делать зарядку читать книги медитировать и рисовать под музыкальное сопровождение < + > сегодня в читаем улыбку рэя брэдбери < + > прямая трансляция в нашем инста а фольтова актриса если вы думаете что вдвоем от того что остальные оркестранты на карантине боже спаси королеву и корону вы ошибаетесь уже сейчас они в четыре руки на одном расстроенном роле прид о дутата состоящего из лидера группы ника бруксовского и сумасшедшего мульти инструменталиста даниила романовича и уже успели подарить эмоции слушателям х городов знакомые и не только мотивы будут в только | тарантино или | спилберг допущены к режиссуре а и каждый желающий может принять участие в создании видеоролика пишите нам а можете даже не писать а сразу брать бабушкину видеокамеру или свой

*Лемматизация (приведение слов к начальной форме):*

```
import pymorphy2

morph = pymorphy2.MorphAnalyzer()

messages = open('posts_processed.txt')
normalized_messages = open('posts_processed_norm.txt', 'w')

for line in messages:
    line_norm = ' '.join(morph.parse(word)[0].normal_form for word in line.split())
    normalized_messages.write(line_norm + '\n')

messages.close()
normalized_messages.close()
,
```

## Фрагмент получившегося массива после лемматизации:

== == == премьера мы решить что хороший способ представить новый сингл сразу снять на он клип клип создать в рамках проект глобальный инициатива бренд | в поддержка артист художник бармен и вес о под этот музыка можно даже поспать но | январь мы точно не удастся поспать мы ожидать новый день без сон который мы повести в новый клуб время билет на концерт тут пока вы допивать последний праз хотя быть чем мы трудиться и очень скоро показать вы новый клип пока всё спасибо за внимание < + > новость съездить в москва переобуться в заодно нахлобучить столичный публика в тонна арбат но как костюм киноман здесь совместно с | мы написать песня к трейлер фильм | майор гром чуной доктор но мы не придумать название дать вариант в комментарий шпана < + > чей название забрать тот презент и б ак бывать мы нужен ваш поддержка всё что накопить на юг за время паломничество отдать воскресный вечером в | не вздумать пропустить событие а то мистер дигидого выловить ты когда быть темно и дать рчество < + > наш состав пополниться новый музыкант и сегодня мы хотеть поделиться второй минифильм создать вместе с в рамках рассказать вчера ребята из что у мы происходить < + > знакомиться прелюд онь потом как то снимать лайва а ник гнать мы из гримёрка выключить камера з тётка дядечка | natalya ashurovskiy | говорить ник переступать с нога на нога ненароком вставать в поза любой античный с заеивать дымить светить таскать крепить | игорь isaakov который помогать весь и по дело мочь бы снять смешной док про | арсений протон тогда ещё обладать полный комплект зуб который копать мёрзлый з торой < + > представлять вы следующий работа цикл минь фильм для альбом режиссёр | в роль | natalya ashurovskiy | даша чернявский при поддержка ленфильм < + > в встречаться на платформа сыграть изоляции вать в течение неделя быть выходить с вы на связь так что следить за новость чтобы не пропустить наш прямой эфир если хотеть провести что то вместе не стесняться писать мы до встреча вечером любимы ко скоро точно новый музыка новый артист новый шляпа у ник сейчас же с превеликий удовольствие похвастаться арт подарок от наш талантливый | акварелька < + > новый дрянн начинать отсчёт новый су тол н < + > небольшой фотчёт о тот как мы посетить столица с новый программа культурный отыграть секретный концерт на день рождение цифербург в день весь влюбить главный где сыграть в бар ровесник и п ы знать что наш гуру звук | день и ночь трудиться над звучание наш музыка не за гора и новый релиз а пока наслаждаться сингл < + > вспомнить забытый рубрика плейлист от наш друг и коллега прийти оч

## Использование алгоритма k-means:

```
#АЛГОРИТМ K-MEANS
posts_norm <- readLines("posts_processed_norm.txt", encoding="UTF-8")

# Создаем корпус из сообщений
posts.corpus <- Corpus(VectorSource(posts_norm))

# Избавляемся от стоп-слов
stop_words <- read.table('stop_words_russian.txt', encoding="UTF-8")
stop_words <- as.vector(stop_words$V1)

posts.corpus <- tm_map(posts.corpus, removeWords, stop_words)

# Удаляем все лишние пробельные символы
posts.corpus <- tm_map(posts.corpus, stripWhitespace)

posts.tdm <- TermDocumentMatrix(posts.corpus)
posts.tdm
posts.tdm.non.sparse <- removeSparseTerms(posts.tdm, 0.90)

# Кластеризация документов
N <- 5
model <- kmeans(posts.tdm.non.sparse, N)

for (cluster in 1:N) {
  cat("cluster ", cluster, ": ", as.vector(names(model$cluster[model$cluster == cluster])), "\n")
}
```

## Построение облака слов:

```
# Построение облака слов
library(wordcloud)
posts.matrix <- as.matrix(posts.tdm.non.sparse)
words.freq <- sort(rowSums(posts.matrix), decreasing = TRUE)
wordcloud(words = names(words.freq), freq = words.freq,
  scale = c(5, 1.0), min.freq = 5, random.order = FALSE,
  rot.per = 0.35, use.r.layout = FALSE, colors = brewer.pal(8, "Dark2"))
```

Получившийся тематический профиль сообщества, представленный в виде облака слов, наглядно демонстрирует, какие темы и ключевые слова преобладают в постах паблика. Проанализировав профиль, можно сделать предположения даже о философии группы. Чем более крупным шрифтом обозначены концепции в облаке слов, тем более часто они встречаются и, следовательно, тем более значимы; мелкий же шрифт, напротив, указывает на меньшую частоту употребления. Разные цвета разделяют концепции на тематические подгруппы.

*Тематический профиль сообщества:*



- Зелёный: слова, относящиеся к основным описаниям постов.
- Синий: инновационно-коллективная составляющая, связанная с идеологией фрипопа, т.е. свободной популярной музыки.
- Оранжевый: идеологическая составляющая – мир, звук, май.  
\*Появившееся в этой группе слово «го», имеющее несколько мусорный оттенок, наводит на мысль о том, что группе было бы правильнее не использовать его в столь больших количествах.
- Бирюзовый: слова, относящиеся ко времени и месту проведения концертов.

Построим граф подписчиков сообщества.

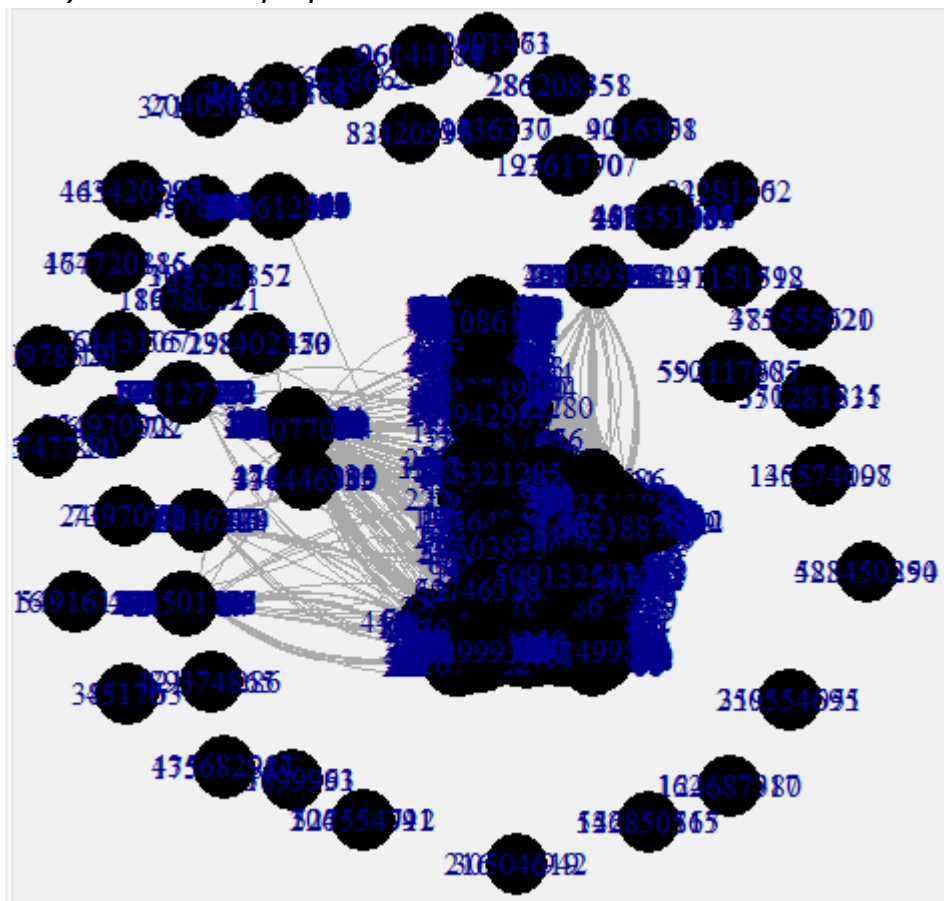
*Вершины – подписчики, рёбра – связи дружбы:*

```
members <- getGroupsMembersExecute('supcolorc')
members_network <- getArbitraryNetwork(members)

g <- graph_from_edgelist(as.matrix(members_network), directed = FALSE)
tkplot(g)
```

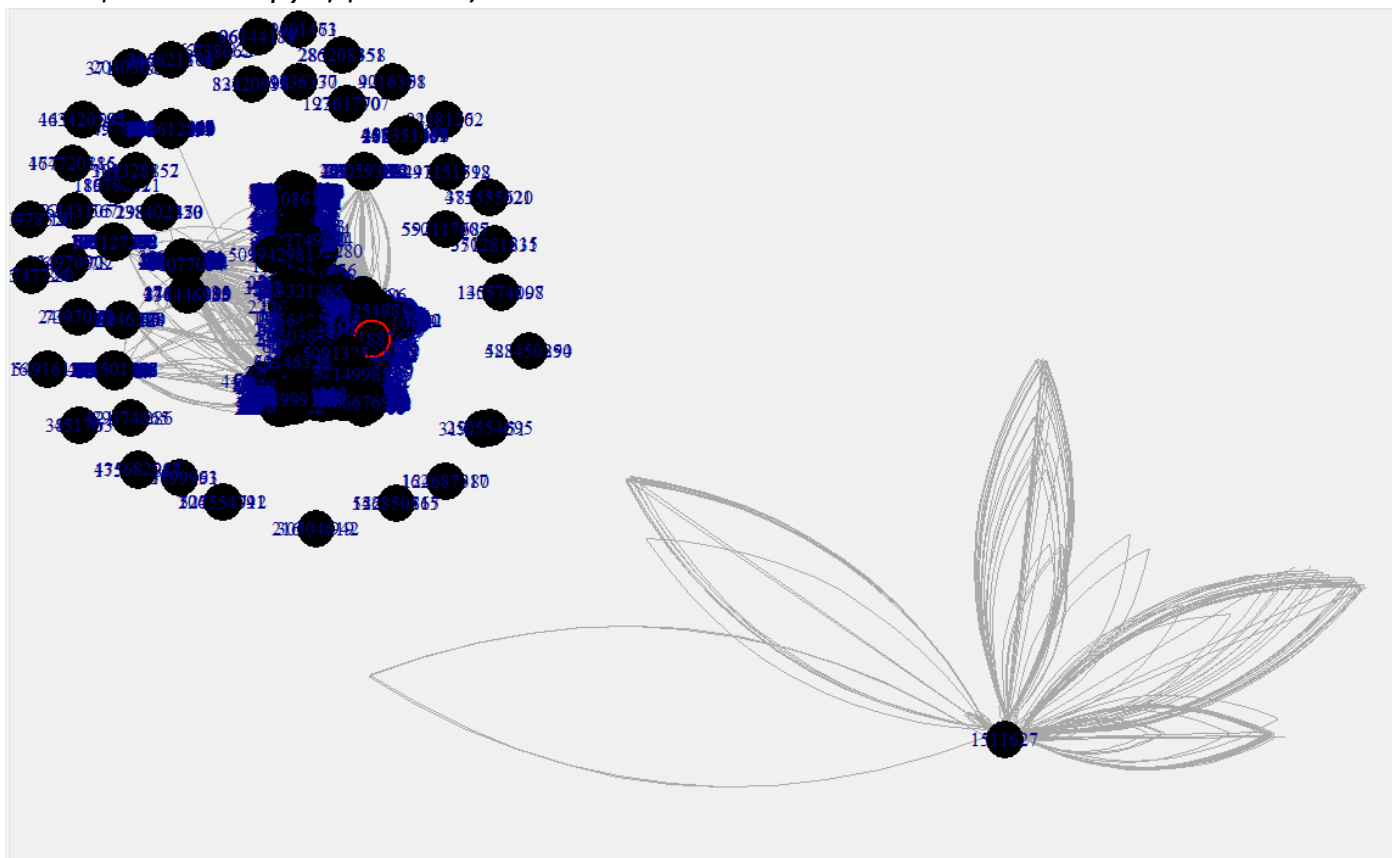


Получившийся граф:



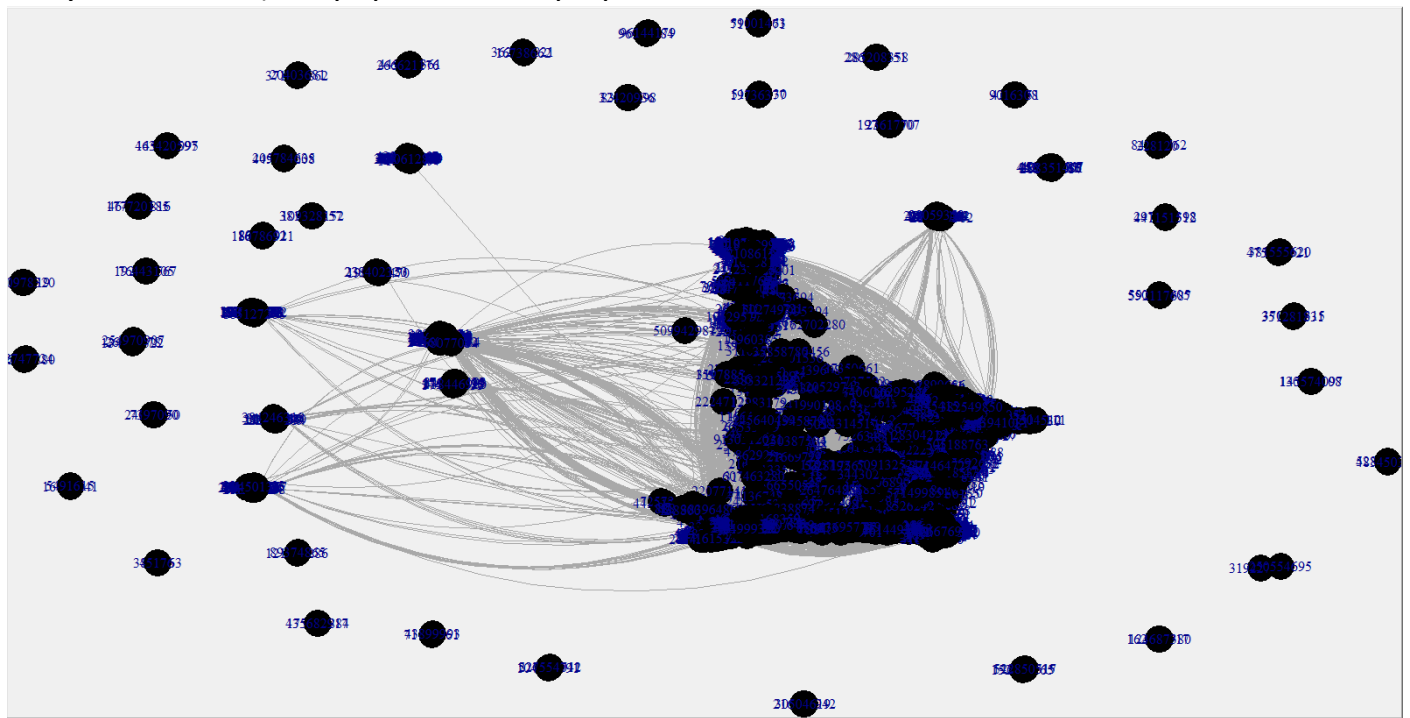
Отцентрируем граф.

Центром графа оказался басист Даниил Романович (справа – центрирование сообщества вокруг Даниила):





Получившийся центрированный граф:



Видно, что граф является связанным не только за счёт наличия звёзд с большим количеством соединений (в нашем случае – музыкантов), но и потому, что многие подписчики сообщества связаны друг с другом. Крайне важно, что в графе присутствуют и изолированные вершины: это говорит о том, что среди подписчиков Super Collection Orchestra не только их друзья и ближайшее окружение, но и незнакомые люди не из их круга, которые заинтересованы непосредственно в музыке.

Количество рёбер в графе:

```
> edges <- nrow(members_network)
> edges
[1] 39487
```

Число связанных компонент (связанная компонента – максимальное число вершин, в котором между любыми вершинами можно проложить путь):

```
> connected_components <- components(g)$no
> connected_components
[1] 39
```

Доля вершин в максимально связанной компоненте:

```
> vertex_giant_component <- max(components(g)$size) / sum(components(g)$size)
> vertex_giant_component
[1] 0.9709203
```

Плотность сети (отношение прямых связей к общему возможному количеству связей) без изолированных подписчиков:

```
> density_without_isolates <- 2 * edges / (vcount(g) * (vcount(g) - 1))
> density_without_isolates
[1] 0.009246452
```

Модулярность (насколько граф поддаётся разбиению на подгруппы):

```
> modularity <- modularity(g, components(g)$membership)
> modularity
[1] 0.00430029
```

Количество кластеров:

```
> clusters <- cluster_louvain(g)
> length(clusters)
[1] 55
```

Средняя геодезическая дистанция (средний кратчайший путь между любой парой узлов в сети):

```
> mean_geodesics <- mean_distance(g)
> mean_geodesics
[1] 3.350568
```

Диаметр сети (расстояние между двумя наиболее удалёнными точками сети):

```
> diameter <- diameter(g)
> diameter
[1] 10
```

Степень центральности сети (степень близости узла к другим узлам):

```
> mean_degree <- mean(degree(g))
> mean_degree
[1] 27.01813
```

Чтобы понять, к какой модели сети можно отнести наш граф, подсчитаем показатели валентностей его вершин (то есть количества связей у каждой вершины). Для этого экспортируем данные `members_network` в csv, вычислим максимальную, минимальную валентность и среднее арифметическое валентности:

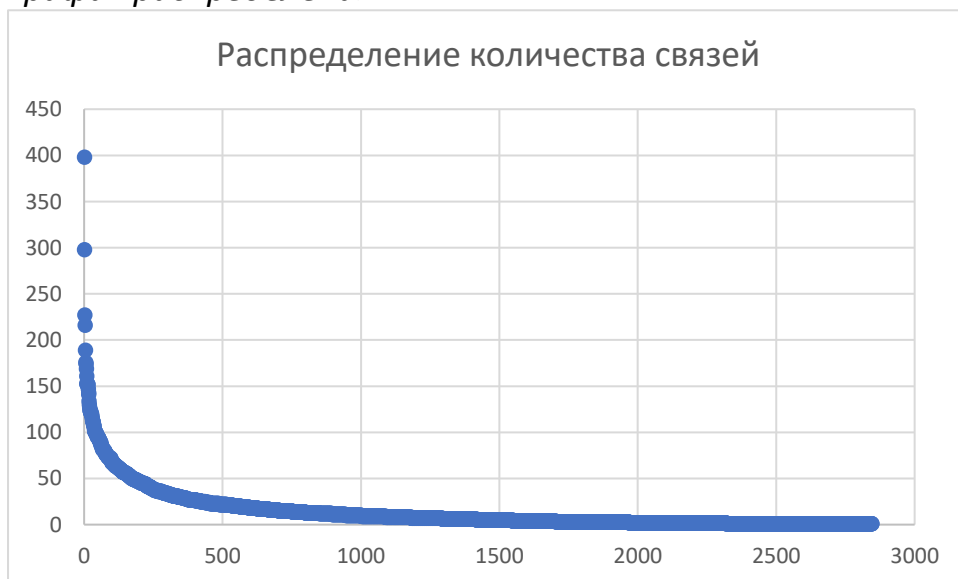
Максимальное значение валентности = 398

Минимальное значение валентности = 1

Среднее арифметическое = 13.8

Максимальная валентность (наибольшее количество связей в графе) обнаружилась у фронтмена группы Ника Брусковского. Басист Даниил Романович на четвёртом месте, а гитарист Андрей Квачёв на 398 месте по количеству связей (их у него всего 27) – это можно объяснить его сравнительно недавним появлением в группе.

График распределения:



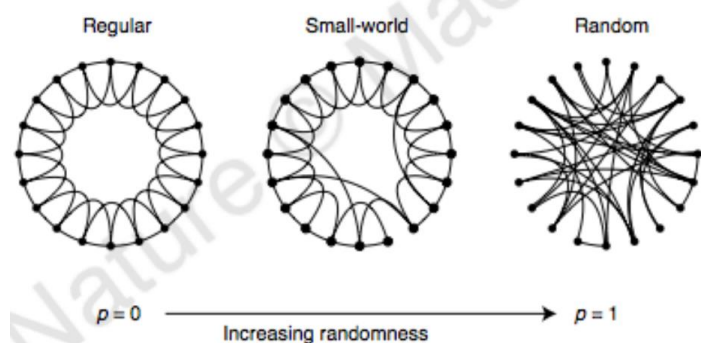
Подсчитаем вероятность модели маленького мира по формуле:

$$P(\text{small world}) = C - L(p) = \ln(\text{MaxValency}) - \ln(\text{Vertices}) - \frac{\ln(\text{MeanValency})}{\ln(\text{MaxValency})}$$

$$= 6 - 8.4 - \frac{2.6}{6} = -2.8$$

Мы получили отрицательную вероятность, значит, этот граф нельзя однозначно отнести к типу small world.

Разберёмся, к какому из типов – small world или случайный граф – относится наша модель:



По форме граф сообщества Super Collection Orchestra можно отнести к циклическому типу, т.к. в нём достаточно велико количество кластеров, а также высока степень центральности.

Сравним граф Small World и случайный граф, чтобы понять, к какому типу отнести нашу модель.

	Small World	Случайный граф	Наша модель
Модель			
Средняя степень вершины	3.833	2.833	13.8
Средняя длина кратчайшего пути	1.803	2.109	3.35
Коэффициент кластеризации	0.522	0.167	0.7

Мы видим, что у графов Small World средняя степень вершины и коэффициент кластеризации выше, чем у случайных, а средняя длина кратчайшего пути между узлами – ниже.

У графа Super Collection Orchestra наблюдается высокий показатель средней длины кратчайшего пути, но при этом достаточно высокие средняя степень вершины и коэффициент кластеризации; это можно посчитать достаточным условием для того, чтобы отнести его к типу Small World.

Тест Колмогорова-Смирнова для проверки гипотезы о принадлежности значений выборки к определённому теоретическому закону распределения:

```
> x <- rnorm(members_network)
> ks.test(x, "pnorm")

      one-sample kolmogorov-smirnov test

data:  x
D = 0.66238, p-value = 0.228
alternative hypothesis: two-sided
```

Тест показывает, что распределение относится к двустороннему типу (симметричное); вероятность ошибки составляет 0.228; расстояние между эмпирической и обыкновенной кумулятивными функциями распределения равно 0.66238.

Вывод: было проанализировано сообщество подписчиков Super Collection Orchestra; выведены его активные пользователи; получен график активности реакций; построен тематический профиль постов сообщества; проведена классификация сети общения.