



Twitter データを利用した 「新型コロナウイルス」関連語句の分析

峰 滝 和 典

要旨 2020年5月21日から30日の期間、「新型コロナウイルス」に関する Twitter のツイートデータを取得して、テキストマイニングを行った。その結果、緊急事態宣言の解除がもたらした影響について Twitter のツイートデータ上どのようにとらえられているかが分かった。経済活動の再開に関するツイートが多い一方、後半以降新たな感染者の発生に関するツイートが既に増加していることがわかった。

Abstract The purpose of this research is to analyze the effects of lifting the state of emergency by collecting data of tweets in Twitter related to COVID-19. The Japanese government lifted it for Osaka, Kyoto and Hyogo on 21 May, for other prefectures on 25 May, 2020. This research was conducted from 21 to 30 May 2020. The results of text mining imply that tweets of the restart of economic activities can be observed through this period, and on other hand tweets which concern the emerge of infected persons by COVID-19 have already increased in the second half of this period.

キーワード Twitter, 新型コロナウイルス, テキストマイニング, 共起ネットワーク, 階層的クラスター分析

原稿受理日 2020年6月1日

1. はじめに

Twitter が社会に大きな影響力を持っているのはこの一カ月^(注1)でも、5月8日「#検察庁法改正案に抗議します」というハッシュタグを付けた抗議が短時間で大量に拡散したケースや、米 Twitter 社が5月29日、暴力を賛美する内容を禁止する利用規定に違反したとして、トランプ大統領によるツイートの上に警告をかぶせ、クリックしないと表示されないようにしたケースから伺える。

Twitter 社はエンドポイントへのアクセスを検討している開発者や研究者向けにスタンダードアローンアプリケーション用申請フォームを作成している。「公益の利益をサポートすること」を目的としたアプリケーションのみがその対象となり、プライバシーや倫理に関する懸念には十分に配慮しなければならない。

Twitter Japan (2020) は、「Twitter は5月12日（米国時間）新型コロナウイルス感染症（以下、COVID-19）に関連する真偽に論議のある内容や誤解を生じる情報を含んだツイートについて、追加のコンテキストや情報を表示する新たなラベルと警告メッセージを導入します」と述べ、内容によって人々を混乱させたり、誤解を生じさせる可能性のあるツイートに対して適切な対応をとるという。これは、2020年3月23日、公式アカウントで、新型コロナウイルス感染症（COVID-19）に関連する誤情報やデマなどが含まれるツイートに削除を要請すると発表したことに続く対応である。

これまで Twitter の API (Application Programming Interface) は、Twitter 社に申請して認められると利用することが可能であった。桐村喬 (2019) によると、API 仕様変更はこれまでもなされてきているという。2018年8月には大幅な API の大幅な仕様変更がなされ、それまで利用可能であった、User streams API や Site streams API が廃止された (桐村喬 (2019))。研究を継続していく上で、Twitter の API 利用環境は変化していくことには注意しなければならない。

本稿を執筆するにあたって著者も Twitter 社に API 利用の申請し応諾されて研究を開始した。Twitter Developer Platform 上で、コンシューマー API キー、コンシューマー API シークレットキー、アクセストークン、アクセストークンシークレットキーを取得した。ツイート取得にはリクエスト制限がかけられている。15分で180リクエストである。2020年5月21日から5月30日まで継続してツイートを取得するためには、スクリプトを動作させるための物理的な環境を整備する必要があつた (桐村喬 (2019))。

新型コロナウイルスと Twitter に関する研究論文はまだ日本では見当たらないが海外では既に発表されている。Kouzy R. et al. (2020) は COVID-19 に関する誤情報がどの程度広がっているかを検証したものである。COVID-19 感染に関する673ツイートのうち、153ツイート (24.8%) が誤情報を含んでおり、107ツイート (17.4%) が未確認情報を含んでいるという。誤情報率が高いのはインフォーマルな個人 / グループのツイートで誤情報率は33.8%であることが述べられている。

Abd-Alrazaq et al. (2020) は約280,000ツイートを収集して、ツイートをトピックスごとに整理して感情分析を行っている。トピックスは12種類あり、中国、アウトブレイク、肉食、生物兵器開発、COVID-19 による死者、COVID-19 に対する恐怖とストレス、旅行の禁止・警告、経済損失、パニック的買占め、人種差別の増加、マスクの着用、検疫対象であり、そのなかでフォロワー数が最も多かったのがアウトブレイク、「いいね」の数が最も多かったものが経済損失、リツイート数が最も多かったものが肉食、そしてリンクシェアが最も多かったものが中国となっている。また感情分析の結果、12トピックスのうち10項目がプラス、COVID-19 による死者と人種差別の増加がマイナスとなっている。

災害時の Twitter 利用に関しては多数の先行研究がある。三浦麻子 (2012) は東日本大震災時の Twitter でのコミュニケーションについて分析している。三浦麻子 (2012) は Twitter のログ分析を通して大災害に際する人間の情報行動の特徴を示した。ユーザは自らの抱く強い不安感情を多く開示し、また自らの得た多くの情報を他者と共有しようとしていたという (三浦麻子 (2012))。

須田剛裕・小嶋和徳・伊藤慶明・石亀昌明・鳥海不二夫 (2013) は東日本大震災時に Twitter に投稿された約 3 億の Tweet データにおいて、リツイート数などを使って情報の拡散状況とデマにつながりやすいキーワードを分析結果を特徴量としたクラスタリングを行うことで、デマ情報かどうかを推定する試みを行みている。高頻度にリツイートしているツイートしているリツイートしているパースト検出によって着目すべきリツイートの絞り込みを行い、リツイートの深さ情報と感情感性を用いてクラスタリングを行い、デマの推定を行っている。

今回の新型コロナウイルスについても Twitter データを用いた分析結果が発表されている。NTT データは、WEB サイト「イマツイ」にて、2月から5月連休明けまでの期間で新型コロナウイルスに関する人々の意識がどのように変化していったかについて、Twitter 全量データから分析している。自粛に関する話題量を調査した結果、2月から4月中旬まで一貫して「自粛すべき」の話題が多かったものの、4月中旬に一度「自粛すべきではな

い」が上回り、4月下旬からGWにかけて「自粛すべきではない」の話題量が増加していったという。

2. テキストマイニングによる Twitter のツイートデータの分析

今回用いる Twitter のツイートデータは、2020年5月21日に大阪府・京都府・兵庫県の緊急事態宣言解除決定の日から開始した。引き続き5月25日は緊急事態宣言が全国において解除された。

緊急事態宣言の解除がTwitter上でどのように取り上げられているのかを見ることが本稿の目的である。

Twitter のツイートデータには、リプライやリツイートのデータも含まれる。自動的にリツイートを行うリツイート bot も含まれる。同じツイートの文章が繰り返し出てきては、あたかもそれに使用されている言葉が Twitter 内の多数意見とみなされてしまう危険性があるので本稿では排除した。その結果、参考表にあるようなツイート数が実際に取得されたものである。総計20万ツイート超である。先述したリクエスト制限のために全数調査とはなっていない。python library の Tweepy を用いてツイート・データを取得している。

ツイートデータ取得後、テキストマイニングを用いるとで、出語数、単語間の関係性を分析した。テキストマイニングとは、テキストを単語やフレーズなどの単位に分割し、それらの出現頻度や共起関係（同時出現）などを集計し、データ解析やデータマイニングの手法で定量的に解析する分析ツールである。共起ネットワークと階層的クラスタ分析を用いて、単語間の関係を可視化した。形態素解析によって文章を単語に分割するわけであるが、本稿ではテキストマイニングのためのフリーソフトの KH Coder によって自動的に行った。

共起ネットワークでは「距離」を「edge」と呼ぶ。edge は線で表現される。抽出語は node と呼ばれ円で表現され、node は edge によって結ばれる。edge の計測には Jaccard 係数を用いた。

階層的クラスタ分析とは、全対象の類似度を計算し、最も類似度の高いものから順次グルーピングし、最終的に1つのクラスタになるまで繰り返す。結びつきの階層構造を明確にする。クラスタの分け方については、ward 法を用いた。

表1から表10が2020年5月21日から5月30日までの出語数上位150を掲載している。5月21日から5月27日間、「経済」が出語数上位20位に入ってる。緊急事態宣言解除によって、

経済活動への関心が高まっていることが伺える。5月25日には、‘安倍’の出語数ランキングが5位に上昇したのも緊急事態宣言解除の記者会見の報道によるところが大きいと思われる。5月27日以降、‘北九州’が、27日13位、28日10位、29日11位、30日6位と急上昇してのいるのは北九州でのクラスター発生が反映されているものと思われる。‘ブルー’；‘インパルス’；‘感謝’の注目度合いが高まっているのは、医療関係者への感謝を示すブルーインパルスの飛行に関心が寄せられているものと推察する。29日に‘亜’と‘塩素’が20位以内に入っているのは、「次亜塩素酸水」の安全性が話題となっていることを示している。30日に‘電通’が20位以内に入ったのは給付金業務が電通に再委託されていたことの報道を受けた結果だと思われる。

図1から図10は期間内の日々の共起ネットワークを掲載している。ネットワークの可視化を目的としているため、最低出語頻度300以上、上位100位の単語に絞って図示している。nodeの大きさは出後頻度の多さを示している。edgeで結ばれているのは、共起関係があることを示している。‘ウイルス’というnodeと‘経済’というnodeが5月21日から23日にかけて、同じネットワークにありかつ距離が近い。21日には、‘安倍’；‘政府’；‘関西’；‘大阪’；‘京都’；‘兵庫’；‘神戸’が比較的密度が高いネットワークを形成している。5月21日には大阪府・京都府・兵庫県で緊急事態宣言解除が決定されており、そのことがネットワーク図にも反映していると考えられる。22日には、‘大阪’；‘新規’；‘患者’；‘病院’がネットワークを形成している。23日には、‘大阪’；‘陽性’；‘患者’；‘神奈川’；‘東京’；‘北海道’がネットワークを形成している。28日には、‘感染’が最大規模のnodeとなり、‘北九州’；‘東京’；‘新た’；‘確認’といったnodeが‘感染’に次ぐ大きさとなり、ネットワークを形成している。22日以降、徐々に感染に関するツイートが増え、ネットワークを形成していることが伺える。

図11から図20は、階層クラスタ分析の結果である。単語と単語の関係性がどのようなクラスタを経て全体に繋がっているのかを示している。

例えば5月30日の階層クラスタ分析（図20）をみると、‘全国’；‘県庁’；‘大都市’；‘所在地’が一つのまとまりとなって比較的少ないステップで全体に繋がっていることがわかる。‘効果’；‘塩素’；‘亜’も同様である。‘医療’；‘機関’；‘患者’；‘病院’；‘新聞’；‘県内’；‘大阪’；‘陽性’；‘新規’；‘ウイルス’；‘東京’；‘ネット’；‘情報’；‘状況’；‘経済’；‘危機’；‘北九州’；‘福岡’も大きなクラスタを形成しており、それぞれ重層的に全体と繋がっている。これらは新型コロナウイルスの感染と経済活動に関する語句が多く含まれると考えられる。

表1 抽出語 上位150 (2020/5/21)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
コロナ	22,133	厚	527	知事	312
新型	21,251	給付	514	流行	312
ウイルス	9,400	必要	505	レモン	310
感染	9,084	子ども	501	韓国	309
対策	2,395	活性	499	新規	306
影響	2,156	患者	493	協力	300
拡大	2,130	企業	474	質問	300
東京	1,705	大阪	469	除去	300
支援	1,688	負担	464	国民	298
日本	1,645	考慮	462	肺炎	298
医療	1,385	界面	461	放送	297
緊急	1,374	学校	459	安倍	296
事態	1,194	効果	455	従事	287
確認	1,136	病院	453	会見	283
宣言	1,106	延期	444	時事	283
対応	1,053	予定	432	専門	283
経済	1,032	研究	421	全国	282
防止	989	多い	414	接触	281
検査	986	開発	413	免疫	279
再開	951	施設	409	抗体	277
情報	929	国内	404	皆様	274
解除	889	消毒	403	日本人	273
新た	847	米	393	学生	272
世界	802	事業	392	少ない	269
営業	799	対象	389	アメリカ	267
中国	796	更新	388	評価	267
死亡	767	神奈川	379	機関	266
関連	757	外出	377	決定	266
マスク	747	利用	374	北海道	265
発表	745	関係	369	安全	259
ワクチン	723	活動	365	継続	259
検討	720	急増	364	リスク	252
状況	693	要請	364	良い	250
中止	691	陽性	363	理解	249
自粛	676	マップ	362	疾患	248
政府	652	社会	359	自分	246
職員	646	詳細	352	動画	246
介護	645	開始	350	特別	246
死者	611	ロード	349	話	244
コロナ	599	県内	349	風邪	243
お願い	580	発生	342	早い	242
有効	578	緩和	338	飲食	241
可能	548	危機	338	調査	241
公開	547	実施	336	変更	241
予防	547	ママ	331	終息	237
生活	542	治療	331	埼玉	235
開催	534	申請	330	英	234
休業	533	相談	319	会議	234
最大	531	臨時	317	配信	234
お知らせ	528	症状	316	サイト	233

表2 抽出語 上位150 (2020/5/22)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	9,400	肺炎	298	ネット	196
東京	1,705	安倍	297	ご覧	192
日本	1,645	時事	283	都内	191
医療	1,385	専門	283	環境	190
コロナ	1,241	全国	282	集団	190
ウイルス	1,108	免疫	279	最新	189
経済	1,032	抗体	277	大学	186
情報	929	皆様	274	段階	186
新た	847	日本人	273	店舗	184
世界	802	学生	272	大会	181
中国	796	少ない	269	期間	180
マスク	747	アメリカ	267	日銀	177
ワクチン	723	機関	266	イベント	176
状況	693	北海道	265	大変	173
新聞	675	安全	259	インフルエンザ	171
政府	652	リスク	252	東洋	168
職員	646	良い	250	体制	167
コロナ	621	疾患	248	電子	165
死者	611	自分	246	個人	164
有効	578	動画	246	家庭	163
可能	548	特別	246	記者	163
最大	531	風邪	243	千葉	163
必要	505	早い	239	委員	161
子ども	501	埼玉	235	業界	160
活性	499	英	234	お客様	159
患者	493	サイト	233	ライブ	156
企業	474	米国	233	政策	155
大阪	469	男性	230	テレビ	154
学校	459	理由	230	無料	154
効果	455	外国	229	詳しい	152
病院	453	無い	229	日常	152
多い	414	自宅	227	家族	151
国内	404	大統領	220	中小	151
事業	392	首相	219	規模	150
米	392	女性	217	公式	150
対象	389	新着	216	現場	149
神奈川	379	最少	215	時点	149
陽性	363	小池	215	政治	148
社会	359	重症	213	独自	148
詳細	351	新しい	210	スーパー	147
県内	349	ブラジル	209	警察	147
危機	338	医師	209	データ	146
新型	318	制度	208	産業	146
臨時	317	怖い	208	強い	145
症状	316	海外	207	明らか	145
緊急	312	不安	206	健康	144
知事	312	センター	203	政権	144
韓国	309	会社	202	洗剤	144
新規	306	高い	202	下記	143
国民	298	地域	200	方法	143

表3 抽出語 上位150 (2020/5/23)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	8,950	最少	329	バス	182
日本	2,582	全国	313	人間	181
東京	1,519	早い	312	時点	180
コロナ	1,419	北海道	306	病床	180
ウィルス	1,331	レンタカー	303	方法	180
マスク	1,157	危機	289	会社	177
世界	1,129	対象	287	子供	177
安倍	1,084	免疫	282	集団	177
中国	1,051	重症	279	男性	177
医療	996	企業	277	皆様	176
国民	956	無い	277	最多	176
経済	941	高い	273	首相	175
活性	908	日本人	264	時事	174
情報	785	女性	262	産業	173
政権	765	埼玉	261	臨時	171
コロナ	748	データ	257	学生	170
政府	747	医師	251	無料	170
理由	743	規模	249	テレビ	167
海外	741	都内	248	沖縄	167
死者	736	マラリア	247	写真	167
状況	702	洗剤	245	外国	166
米	655	子ども	244	厚生	166
効果	646	ネット	243	お客様	164
患者	642	新着	240	大きい	164
メディア	637	事業	239	自宅	163
ワクチン	623	知事	239	欧米	159
新聞	597	武漢	228	アルゼンチン	157
大阪	593	地域	227	センター	157
新規	492	大統領	223	市長	156
社会	478	韓国	221	労働省	156
多い	468	機関	221	ライブ	152
良い	468	抗体	220	中心	150
病院	458	家庭	219	千葉	149
学校	445	風邪	219	先生	146
新型	429	新しい	216	府知事	146
ブラジル	427	米国	215	キャンペーン	145
陽性	426	サイト	213	低い	145
症状	403	個人	211	悪い	144
少ない	385	全面	211	現場	143
県内	384	怖い	211	自転車	142
国内	371	トランプ	204	南米	142
自分	371	総理	204	感じ	141
最大	369	動画	202	段階	141
専門	368	最新	200	状態	139
神奈川	359	政治	198	内閣	139
職員	348	インフルエンザ	197	週間	138
肺炎	345	債務	189	責任	138
リスク	339	大手	189	被害	138
アメリカ	335	都市	187	衛生	137
期間	332	新興	186	WHO	136

表 4 抽出語 上位150 (2020/5/24)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	11,576	武漢	427	香港	242
東京	3,159	都内	425	ネット	241
日本	2,333	学校	422	研究所	236
コロナ	2,175	県内	409	お客様	232
マスク	2,163	症状	390	匿名	228
政府	1,652	方針	389	データ	227
ウィルス	1,550	首相	365	低い	222
医療	1,456	少ない	365	院内	217
経済	1,213	専門	365	皆様	217
世界	1,189	地域	362	女性	217
状況	1,158	最大	349	環境	215
中国	1,136	動画	349	京都	210
北海道	1,102	経路	340	男性	207
死者	1,099	医師	337	外国	206
安倍	944	アメリカ	332	札幌	205
全面	942	高齢	326	知事	202
患者	921	全国	325	臨時	202
情報	844	ブラジル	312	臨床	202
コロナ	808	重症	312	大きい	201
良い	798	日本人	312	時点	199
多い	771	都県	311	制度	198
クロ	747	無い	307	都市	198
新聞	719	インフルエンザ	303	細胞	197
ワクチン	653	千葉	300	現場	194
社会	637	新着	296	世界中	194
国民	622	埼玉	294	イベント	193
新規	612	風邪	291	学生	192
陽性	609	危機	289	人間	192
大阪	589	会社	287	静岡	192
米	559	時事	287	課題	191
病院	554	福岡	287	職員	191
高い	538	厚生	281	店舗	188
新型	529	個人	280	委員	186
海外	517	サイト	278	状態	185
通気	516	韓国	278	悪い	183
素材	505	怖い	278	大学	183
首都	499	米国	275	段階	183
政権	497	期間	272	大統領	179
効果	494	総理	270	様式	179
早い	489	事業	268	場所	178
機関	479	対象	264	原因	177
神奈川	469	新しい	263	北九州	177
リスク	467	政治	262	方法	175
国内	463	本部	261	各国	174
メディア	457	最新	257	写真	174
免疫	450	労働省	252	センター	173
自分	446	集団	249	子ども	173
肺炎	431	抗体	248	イン	172
企業	429	自宅	247	子供	172
理由	429	テレビ	244	保険	171

表5 抽出語 上位150 (2020/5/25)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	13,166	神奈川	426	個人	288
日本	3,381	新着	417	危険	286
コロナ	2,099	早い	417	都内	286
東京	1,974	専門	416	日本人	286
安倍	1,608	結核	415	委員	285
ウイルス	1,570	歴史	414	不安	279
政府	1,501	リスク	411	段階	276
マスク	1,479	理由	409	晋	274
首相	1,399	新規	407	最大	272
医療	1,324	対象	406	子ども	272
世界	1,279	特別	406	モデル	271
中国	1,240	学校	402	イベント	270
情報	1,233	陽性	395	学生	269
経済	1,190	機関	393	世界中	265
死者	1,124	新しい	390	怖い	265
米	1,010	症状	385	医師	258
新た	968	闘い	382	現場	258
ワクチン	950	県内	381	スポーツ	257
状況	945	戦略	381	規模	256
可能	770	韓国	374	基準	248
必要	761	野球	368	大臣	246
全国	760	サイト	364	風邪	246
コロナ	724	皆様	359	武漢	245
社会	714	高い	358	データ	242
患者	703	米国	355	責任	241
国民	687	背景	353	千葉	240
事業	673	臨時	346	予算	240
全面	656	無い	345	免疫	239
多い	652	インフルエンザ	340	有効	237
新聞	613	地域	340	首都	236
企業	569	プロ	339	大会	235
北海道	543	記者	339	知事	235
抗体	537	安全	337	大変	233
政権	523	最新	337	様式	232
国内	521	西村	337	時点	230
海外	519	アメリカ	335	治験	230
緊急	502	動画	334	北九州	230
人工	498	期間	331	お客様	227
自分	497	ブラジル	330	健康	222
病院	490	会社	324	ライブハウス	217
総理	474	重症	322	皆さん	217
保健所	474	メディア	317	時事	217
肺炎	473	危機	316	課題	216
大阪	468	効果	307	不要	215
良い	460	詳細	307	店舗	208
新型	456	政治	307	時代	207
本部	442	ネット	304	各国	206
方針	436	テレビ	302	外国	205
独自	432	埼玉	298	国際	204
少ない	431	福岡	289	市内	204

表6 抽出語 上位150 (2020/5/26)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	13,191	抗体	393	知事	250
日本	4,019	危機	375	戦略	248
東京	1,716	対象	370	個人	247
中国	1,661	皆様	362	埼玉	246
コロナ	1,606	新規	361	新報	245
ウイルス	1,502	肺炎	356	治験	243
マスク	1,442	会社	354	匿名	242
医療	1,437	サイト	351	アルコール	239
世界	1,367	米国	343	高齢	239
安倍	1,277	免疫	342	時代	239
経済	1,264	インフルエンザ	339	福岡	239
情報	1,139	機関	337	電子	237
死者	1,138	韓国	335	ネット	235
政府	1,116	神奈川	335	医師	234
首相	1,001	北海道	333	子ども	234
状況	891	期間	332	方針	233
ワクチン	864	理由	332	ビル	231
米	771	事務	331	ゲイツ	230
多い	768	県内	328	クイズ	229
企業	752	政権	325	医事	229
コロナ	709	早い	324	重症	229
国民	699	WHO	323	時事	228
患者	648	アジア	322	店舗	226
新聞	636	予算	322	背景	225
全国	607	高い	321	イベント	224
社会	597	大学	315	悪い	224
事業	550	動画	315	怖い	221
ブラジル	512	現場	306	結核	219
学校	509	パチンコ	301	国際	213
国内	502	無い	301	風邪	213
病院	469	武漢	300	打撃	212
海外	468	局長	297	厚生	209
少ない	466	全面	293	集団	208
政治	463	記者	292	闘い	207
新しい	456	新着	288	内容	207
効果	455	最新	287	家庭	206
アメリカ	454	歴史	287	政策	206
専門	451	保険	286	殺人	205
新型	446	様式	280	体制	205
臨時	437	データ	278	大会	203
総理	433	地域	276	女性	201
北九州	428	日本人	275	市内	199
大阪	423	課題	274	通常	198
症状	410	お客様	268	ホームページ	197
陽性	407	テレビ	266	時点	197
リスク	405	欧米	266	最大	195
メディア	402	段階	258	学生	194
保健所	401	自宅	255	ドット	193
良い	398	陰謀	252	大臣	191
自分	397	人工	252	詳しい	189

表7 抽出語 上位150 (2020/5/27)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	13,627	大村	390	神奈川	254
日本	2,672	危機	384	高い	251
コロナ	2,087	専門	379	福岡	250
東京	1,908	武漢	378	動物	248
情報	1,679	症状	377	インフルエンザ	246
ウイルス	1,491	パチンコ	373	国際	245
医療	1,455	リスク	365	店舗	245
政府	1,294	対象	355	重症	244
マスク	1,291	地域	353	経路	241
中国	1,280	サイト	350	都知事	240
予算	1,175	海外	348	市民	239
世界	1,162	皆様	347	自治体	239
北九州	1,073	ネット	343	スーパー	238
経済	1,068	機関	343	家族	237
死者	1,052	最大	342	男性	235
大阪	1,031	新しい	341	理由	235
職員	930	良い	341	方針	231
状況	890	日本人	340	詳しい	229
安倍	798	メディア	339	スポーツ	228
陽性	733	集団	334	学生	228
患者	719	抗体	331	怖い	226
コロナ	708	総額	331	子ども	223
韓国	670	政治	328	無料	223
新聞	667	米国	327	京都	222
兵庫	666	効果	322	ブラジル	216
国民	665	早い	317	大会	216
加西	662	規模	316	一般	212
企業	661	テレビ	313	需要	212
事業	655	データ	309	小池	212
知事	652	政権	304	総理	208
米	634	動画	300	電子	208
社会	593	センター	299	時点	207
多い	591	市内	297	自宅	207
医師	582	少ない	294	シティ	206
前提	560	政策	293	イベント	205
ワクチン	527	最新	290	災害	198
病院	494	女性	288	福島	198
学校	482	現場	287	年度	196
アメリカ	476	免疫	286	内容	193
国内	470	歳出	282	トランプ	192
臨時	442	個人	278	千葉	191
首相	428	愛知	275	埼玉	190
新型	425	市長	275	地方	189
新規	418	お客様	273	外国	188
会社	411	期間	270	被害	188
全国	404	無い	266	大手	187
県内	402	北海道	265	事態	183
自分	393	業界	263	人口	183
都内	393	大学	260	方法	183
肺炎	391	首都	256	市場	182

表 8 抽出語 上位150 (2020/5/28)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
感染	13,250	中止	574	時事	343
ウイルス	11,790	関連	541	センター	340
対策	3,182	生活	536	自分	334
確認	2,834	休業	531	愛知	333
東京	2,414	新聞	528	効果	327
拡大	2,225	施設	520	高校生	324
新た	2,141	開催	518	重症	309
影響	2,069	県内	505	危機	308
日本	1,924	安倍	498	緊急	308
北九州	1,699	実施	487	増加	306
マスク	1,447	関係	485	神奈川	299
コロナ	1,445	予定	473	コメント	298
医療	1,442	議事	469	ブルー	297
再開	1,325	必要	463	集団	297
ウイルス	1,312	多い	452	検討	295
情報	1,297	治療	451	予算	292
防止	1,198	国内	450	期間	291
対応	1,183	米	449	検証	290
解除	1,181	従事	447	米国	290
死亡	1,148	決定	445	研究	289
検査	1,076	新規	439	着用	289
支援	1,016	延期	438	皆様	288
宣言	1,007	国民	436	制限	288
世界	1,004	新型	436	早い	287
死者	961	都内	432	話	286
営業	948	福岡	425	詳細	281
発表	946	全国	424	崩壊	278
患者	933	利用	405	経路	277
自粛	901	ワクチン	403	良い	277
大阪	880	質問	399	導入	276
状況	879	感謝	394	臨時	275
政府	877	首都	392	海外	273
韓国	827	企業	390	ネット	271
可能	815	症状	390	開始	271
経済	811	接触	384	国家	271
コロナ	787	活動	382	相談	271
専門	779	公開	378	インパルス	270
中国	764	社会	376	政権	270
会議	763	安全	375	短縮	270
連続	729	北海道	374	無い	270
予防	704	肺炎	373	サイト	266
病院	698	速報	371	首相	264
要請	658	流行	368	看護	262
発生	652	収束	367	変更	262
給付	628	協力	365	報道	261
職員	626	学校	364	リスク	260
知事	600	作成	359	減少	260
アメリカ	587	調査	352	市民	260
陽性	584	事業	348	ソウル	259
外出	577	消毒	345	機関	259

表9 抽出語 上位150 (2020/5/29)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
感染	12,072	安倍	632	最大	357
ウイルス	12,007	効果	616	肺炎	357
確認	3,227	官房	615	都内	355
対策	3,165	死亡	598	国内	354
東京	2,980	新規	593	会見	353
医療	2,766	給付	592	社会	352
対応	2,411	大阪	572	流行	352
拡大	2,213	開催	560	国民	351
影響	2,128	休業	554	現時点	347
新た	2,120	新聞	549	方々	343
北九州	1,961	皆様	544	福岡	330
ブルー	1,899	関連	533	学校	329
インパルス	1,868	連続	528	機構	328
会議	1,840	可能	524	トランプ	320
専門	1,787	自衛隊	519	開発	318
感謝	1,766	評価	511	増加	318
従事	1,690	予防	508	協力	317
塩素	1,564	開示	504	発言	317
日本	1,556	知事	495	労働	317
亜	1,552	予定	491	首相	314
コロナ	1,497	生活	488	決定	313
飛行	1,469	病院	487	会社	312
ウィルス	1,296	多い	483	外出	312
再開	1,276	都心	483	サイト	308
情報	1,205	必要	474	開始	308
有効	1,163	韓国	473	早い	308
宣言	1,161	活動	458	機関	306
防止	1,150	ブラジル	457	目的	306
敬意	1,147	長官	456	基盤	305
検査	1,111	発生	453	議論	304
議事	1,092	航空	442	記録	303
経済	1,020	実施	436	危機	301
マスク	1,006	企業	428	公開	301
発表	947	死者	426	速報	301
政府	942	市長	419	研究	298
営業	923	施設	419	事業	298
支援	883	新型	416	内閣	294
世界	862	米	397	リスク	293
状況	835	要請	394	安全	292
関係	822	ワクチン	391	変更	292
調査	817	延期	391	緊急	291
解除	800	技術	387	自由	290
中国	794	陽性	387	政権	288
消毒	771	全国	385	自分	286
コロナ	727	治療	377	販売	286
作成	711	良い	374	気持ち	285
患者	674	製品	367	話	284
中止	670	上空	366	時事	283
利用	663	アメリカ	363	無い	280
自粛	643	報道	359	悪化	279

表10 抽出語 上位150 (2020/5/30)

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
ウイルス	11,228	所在地	321	知事	198
東京	1,975	福岡	318	個人	197
日本	1,536	国民	317	免疫	196
医療	1,433	機関	316	現場	189
コロナ	1,397	政権	316	現時点	186
北九州	1,298	肺炎	314	地域	186
マスク	1,228	海外	311	センター	183
ウィルス	1,205	自分	310	テレビ	183
世界	928	リスク	302	陰性	183
経済	909	症状	302	時点	180
専門	896	首相	299	制度	180
中国	849	早い	298	武漢	180
政府	843	皆様	294	お金	178
情報	842	社会	294	時事	178
状況	780	無い	278	怖い	178
議事	769	随意	275	沖縄	176
業務	724	韓国	274	風邪	173
米	709	ブルー	272	市民	172
電通	691	ブラジル	264	対象	168
全国	646	無料	259	イベント	167
患者	629	千葉	258	メトロ	167
コロナ	626	女性	254	岡山	167
陽性	556	システム	253	概要	167
安倍	553	インパルス	251	通常	167
死者	546	神奈川	247	政治	166
多い	525	北海道	245	賞与	164
トランプ	523	都内	244	大会	164
新聞	486	お客様	240	政策	163
病院	480	サイト	238	厚生	160
大阪	451	理由	237	家族	159
塩素	449	重症	233	経路	156
亜	447	埼玉	232	ホーム	155
半数	420	新しい	228	時代	155
国内	405	最新	226	低い	154
企業	400	米国	226	予算	154
効果	397	インフルエンザ	221	団体	152
大統領	394	愛知	220	国際	151
ネット	390	医師	220	子ども	151
危機	387	期間	219	状態	150
新規	378	高い	217	都心	150
アメリカ	356	映画	216	強い	149
ワクチン	351	男性	214	西村	149
新型	351	臨時	214	大きい	149
少ない	345	事業	213	労働省	148
大都市	342	抗体	212	人間	147
県内	338	アルコール	203	日本人	147
県庁	335	技術	202	大学	146
学校	330	動画	201	方針	145
良い	327	日	199	データ	144
会社	323	集団	198	当局	144

図3 共起ネットワーク（2020/5/23）

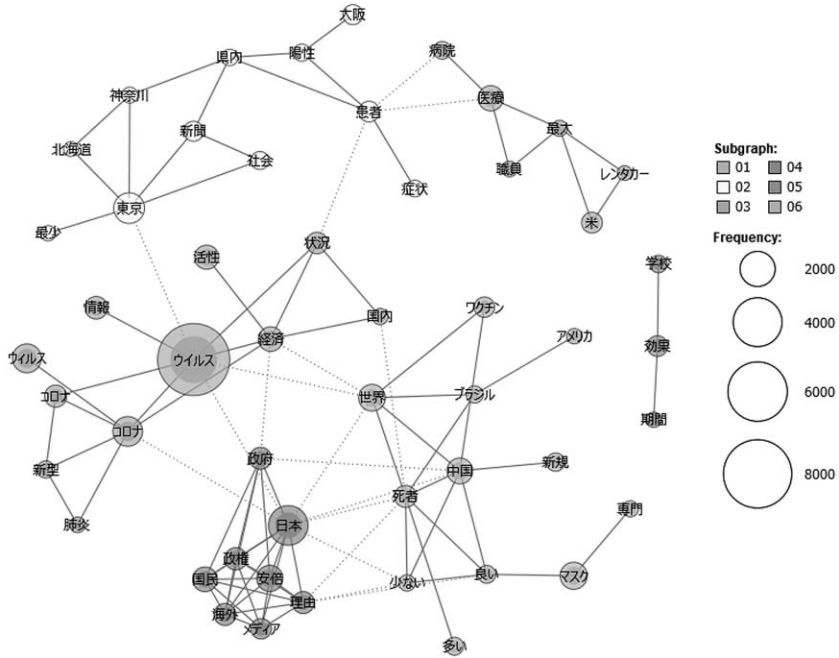


図4 共起ネットワーク（2020/5/24）

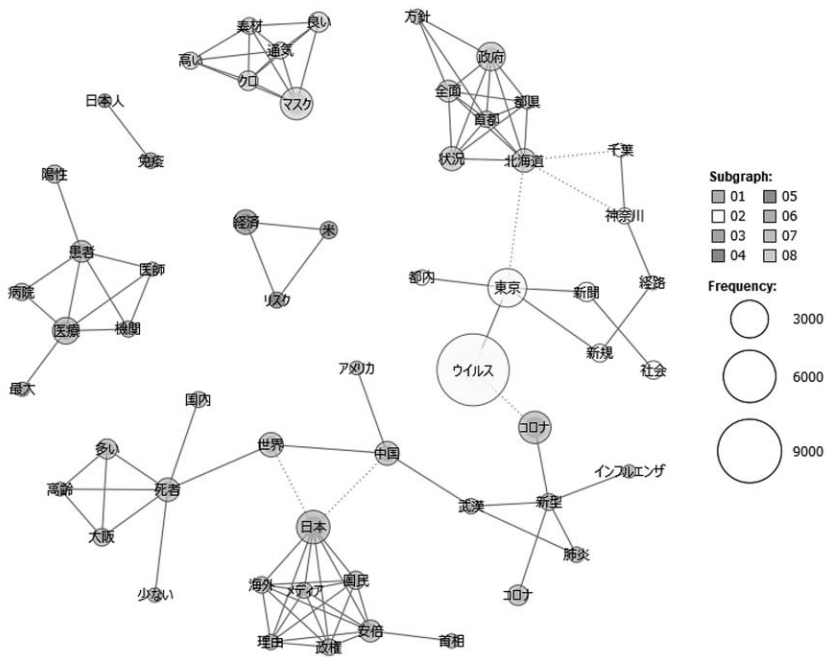


図5 共起ネットワーク (2020/5/25)

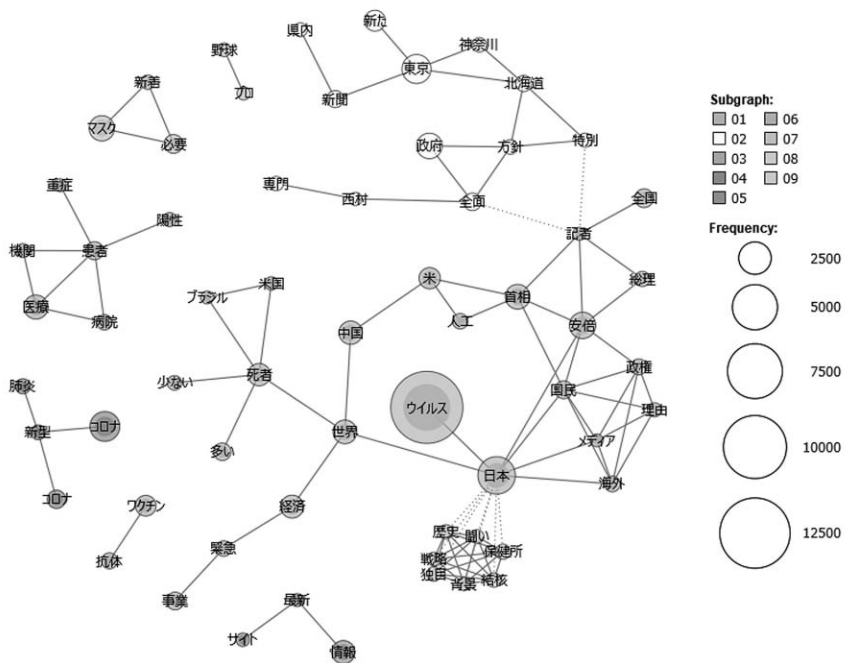


図6 共起ネットワーク (2020/5/26)

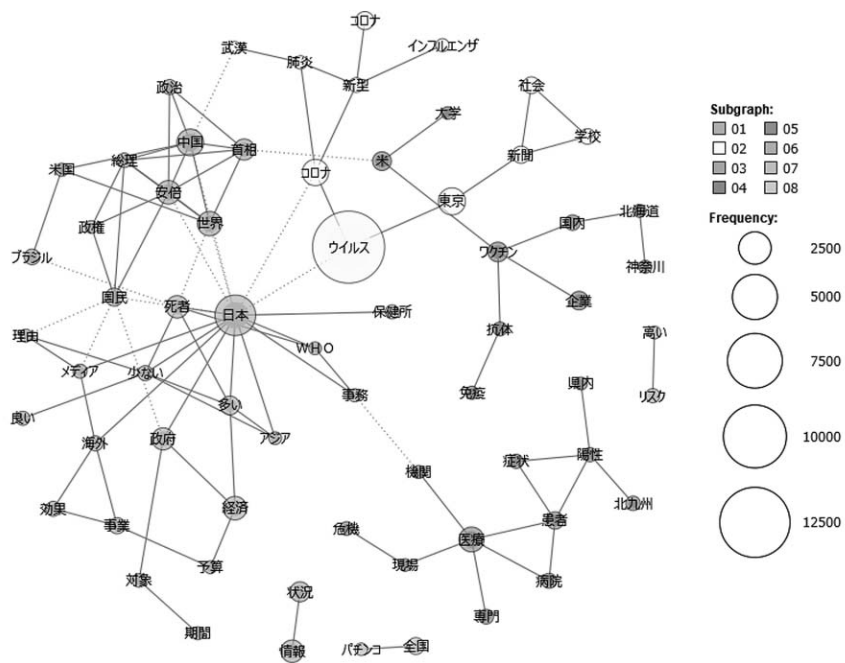


図7 共起ネットワーク（2020/5/27）

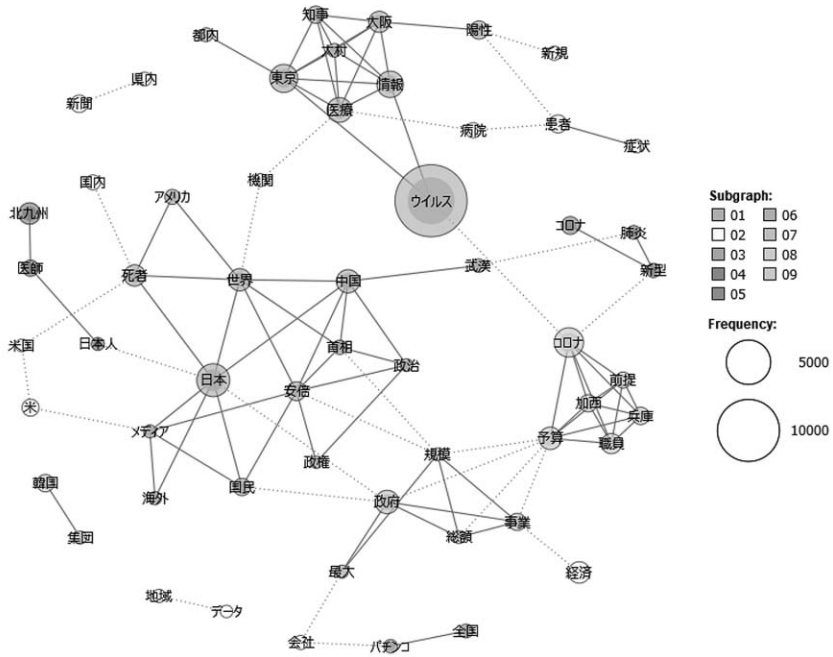


図8 共起ネットワーク（2020/5/28）

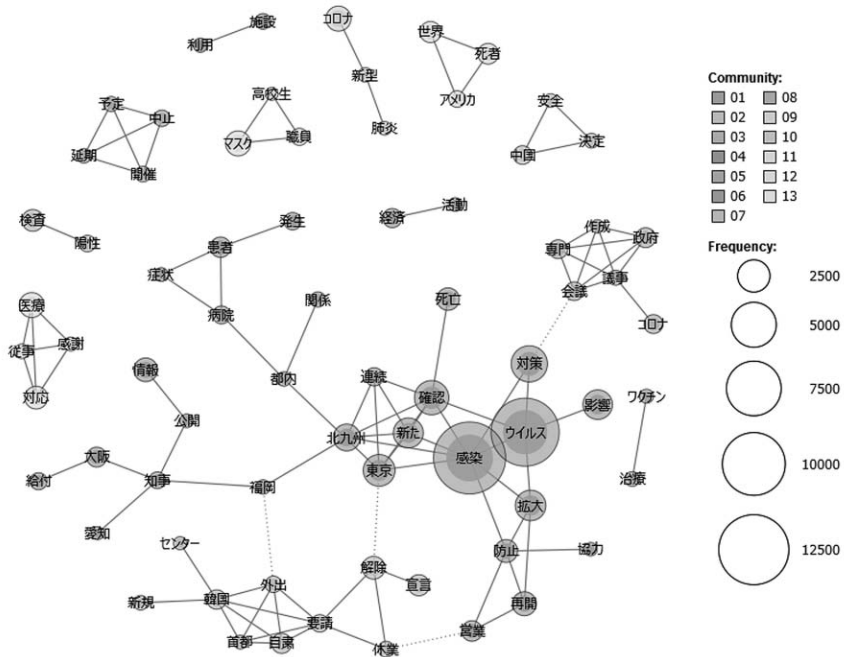


図9 共起ネットワーク (2020/5/29)

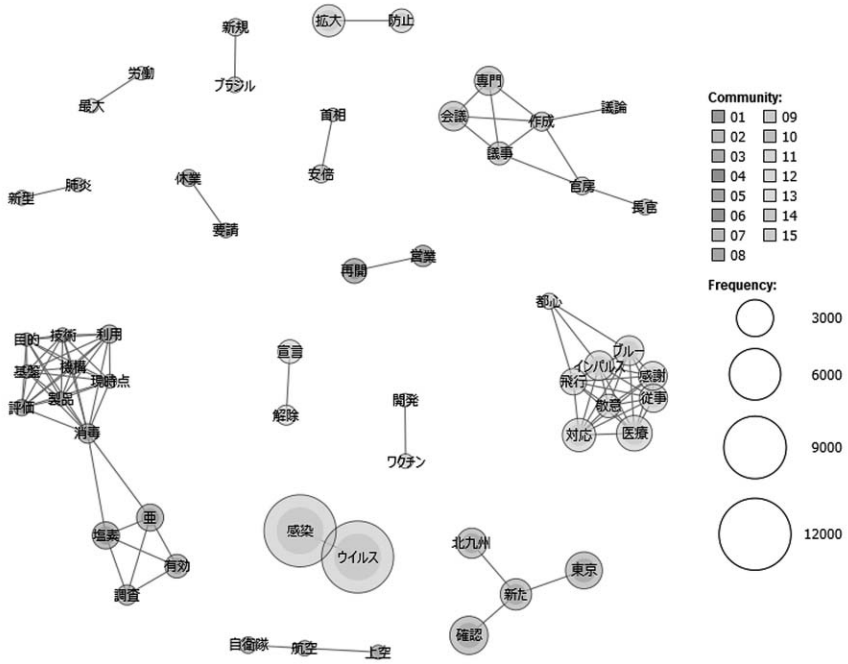


図10 共起ネットワーク (2020/5/30)

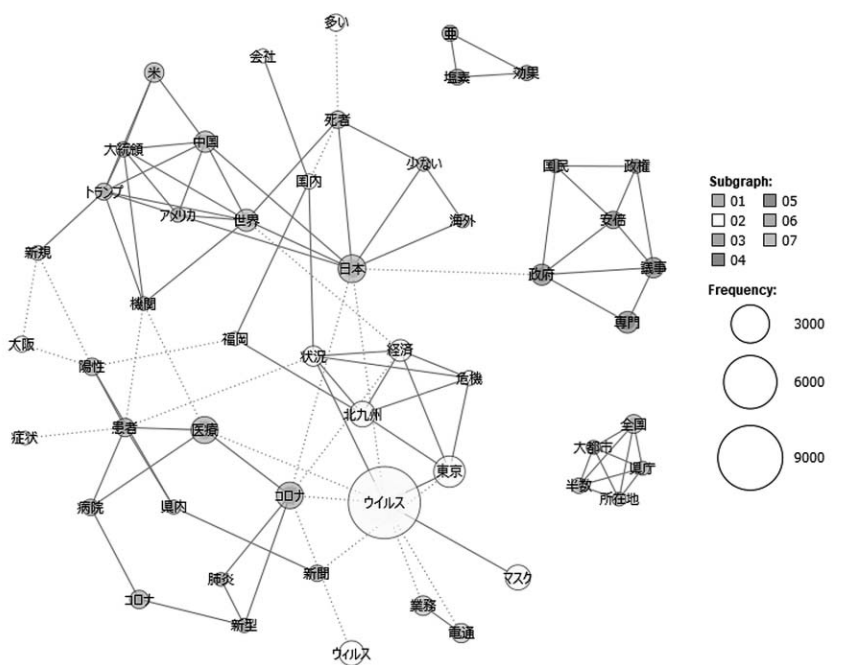


図18 階層的クラスタ分析（2020/5/28）

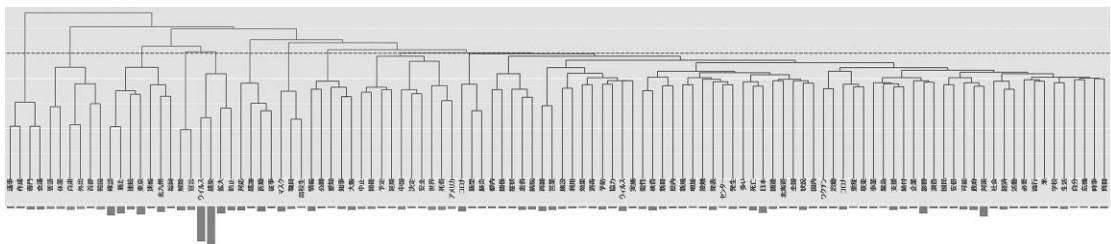


図19 階層的クラスタ分析（2020/5/29）

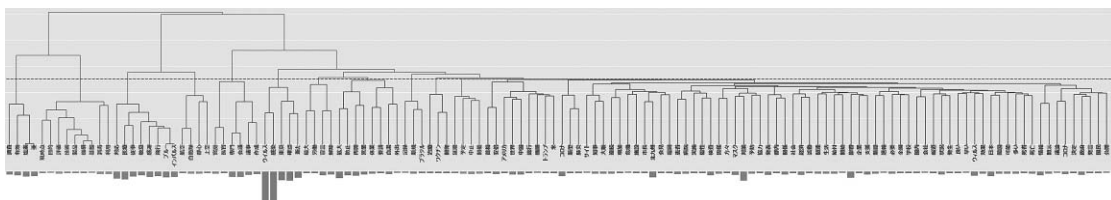
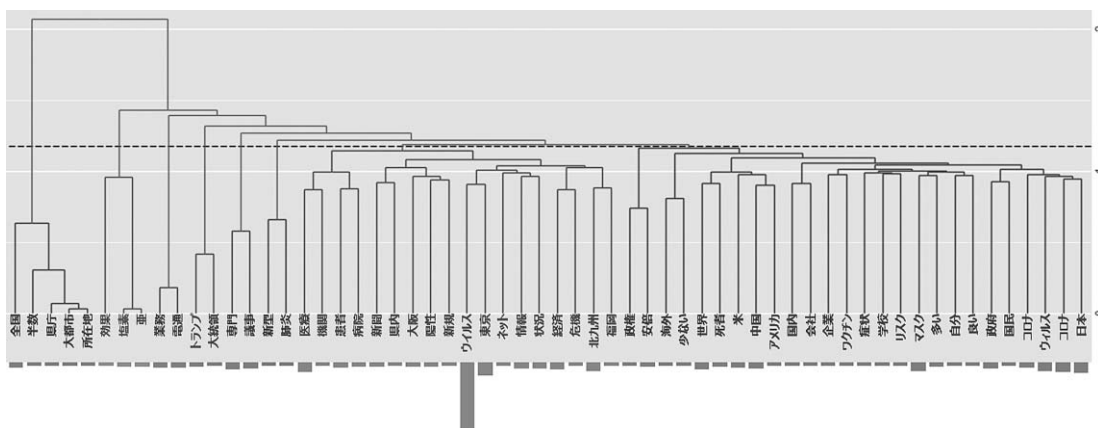


図20 階層的クラスタ分析（2020/5/30）



3. 結び：今後の課題

抽出語の分析、共起ネットワーク、階層的クラスタ分析を行った結果、Twitter のツイート上、緊急事態宣言解除後、経済活動の再開に対する期待を感じる反面、新たな感染者の出現に関する関心の高まりが観測できた。

本稿のテキストマイニングでは、抽出語分析においては単純にツイートのなかに表れる新型コロナウイルスに関連する語句の頻度がわかる。共起ネットワークでは新型コロナ

ウィルスに関するツイート上で、どの語句とどの語句が同時に用いられる頻度が高いかということがネットワーク図で可視化される。階層クラスター分析では、ネットワーク図で表される新型コロナウイルス関連語句のまとまりがどの程度の階層を経て全体に繋がるのかが可視化される。各分析とも時系列の推移をみることで、新型コロナウイルスに関する話題の変遷がわかる。

本稿は、「新型コロナウイルス」に関するツイート分析の準備段階に書かれたものである。今後継続して、Twitter からツイートを継続的に取得して、分析を深めたい。それと同時に、KH Coder によるテキストマイニング以外に、pythonを用いて、MeCabなど他の形態素解析を試みることで今回の結果を検証したいと考える。

また、ツイートをカテゴリー別に分けて感性分析を行なうことも今後の課題としたい。

謝 辞

本稿の分析では樋口耕一氏が作成された KH Coder を用いてテキストマイニングを行った。感謝の意を示したい。

(注1) 本稿執筆時点の2020年5月下旬。

参 考 文 献

- 須田剛裕・小嶋和徳・伊藤慶明・石亀昌明・鳥海不二夫 (2013) 「震災時におけるツイッターのトレンドワードと拡散情報を利用したデマ推定の一考察」 情報処理学会第75回全国大会
- 桐村喬 (2019) 『ツイッターの空間分析』(古今書院)
- 樋口耕一 (2020) 『社会調査のための計量テキスト分析 ―内容分析の継承と発展を目指して― 第2版』, ナカニシヤ出版
- 三浦麻子 (2012) 「東日本大震災とオンラインコミュニケーションの社会心理学 ―そのときツイッターでは何が起こったか―」 電子情報通信学会誌 VOL.95 No.3, 2012
- Abd-Alrazaq Alaa, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, Zubair Shah, “Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study”, Journal of Medical Internet Research, 2020 vol. 22 iss. 4
- Kouzy Ramez, Joseph Abi Jaoude, Afif Kraitem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, Khalil Baddour (2020), “Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter”, Cureus. 2020 March
- Twitter Japan (2020), “COVID-19に関して誤解を与える情報への取り組みをアップデート” https://blog.twitter.com/ja_jp/topics/company/2020/Covid19Misinfo.html

Twitter データを利用した「新型コロナウイルス」関連語句の分析（峰滝）

参考表 2020年5月21日から同年5月30日までに、“新型コロナ”というワードで検索したツイート取得数（リプライ，リツイートは除く）

日付	ツイート取得数
5月21日	26,626
5月22日	20,501
5月23日	21,648
5月24日	27,422
5月25日	28,956
5月26日	27,951
5月27日	27,991
5月28日	24,139
5月29日	25,220
5月30日	22,436