

テキストマイニングを用いた災害情報の抽出と分析

～地方紙 HP が熊本地震をどう伝えたか～

○木原 大志*¹ 小立 雄大*¹
緒方 雄基*¹ 柳 信栄*¹
小林 祐司*²

キーワード：災害 テキストマイニング 新聞 ホームページ

1. はじめに

2016 年 4 月 14 日に発生した熊本地震では、熊本県を震源とし、九州全体に被害が広がった。一連の地震による被害は死者 98 名、軽・重傷者 2,321 名であり、地震活動において震度 7 が二度観測されたのは初めてであった。また、国内の震度 7 の観測事例としては、4・5 例目である。¹⁾

地震の多発する我が国では、災害時のメディアの活用は非常に重要であり、他国に比べ進歩している。例として、緊急地震速報や SNS による安否確認が挙げられる。熊本地震においても携帯電話、テレビ、ラジオを通して被害状況、避難所情報、救援物資情報など多くの情報を得た人が非常に多いだろう。しかし、メディアによって発信された情報の偏りや本当に望まれている情報なのか。また、時間経過による発信情報の変化などを分析した例は少ない。

メディアが発信した情報を分析するツールとして「テキストマイニング (Text Mining)」がある。テキストマイニングは、形式化されていない文章に含まれる単語や文節から、それらの出現頻度、関係性を定量的に把握する手法である。この手法を用いて、小林ら²⁾ はアンケート調査の自由回答を対象として、生活環境評価の地域別傾向や課題を把握し、評価構造を明らかにしている。一方、伊藤ら³⁾ は東日本大震災に関する web アンケート調査から、首都圏での震災時帰宅立ち寄り行動の実証研究を行い、立ち寄り施設と心理・状況の関係を分析した。

これらの分析は、アンケート調査などで被災者から得られた非形式データを基に分析している。しかし、テレビや新聞、ラジオといったメディアが発信するテキストデータの情報について分析した例は少ない。また、災害時の情報発信者として、何をいつ伝えたのかといった情報の整理は重要な分析であるとする。

そこで、本研究では「新聞」に掲載された地震関連記事を対象として、地方紙 HP の記事をテキストマイニング分析の対象とし、地方紙が

どう熊本地震を伝えたのか、また、記事の中に隠れている潜在的な情報を抽出し分析を行う。新聞は膨大な情報を記載することができ、紙媒体として情報を得ることが可能である。また、どのような情報をいつ発信したのかがわかりやすいことから、本研究の対象とした。

2. 分析対象と分析の流れについて

2. 1 分析対象

本研究では、大分県の地方紙である「大分合同新聞社」が運営する「大分プレミアムオンライン Gate」⁴⁾ を分析対象の地方紙 HP とする。熊本地震発生後、地震速報や余震情報、大分県内外の被害状況、避難所情報などを随時更新し、紙面を購読していない人もインターネットを通して閲覧可能である。

分析対象の記事は、地震発生 (2016 年 4 月 14 日) から 2016 年 6 月 15 日までの二か月間に掲載された地震関連記事とする。地震情報、被災地情報、速報、避難所情報、地震によるイベント延期情報なども対象とした。記事の収集については複数人で行い、表 1 に記事の一部を示す。

2. 2 分析の流れ

本研究では、KH coder^{注 1)} を分析ツールとして利用する。KH Coder とは、テキストマイニングを行うためのフリーソフトウェアで、アンケートの自由記述・インタビュー記録・新聞記事など、さまざまな社会調査データを分析することができる。分析の流れを Fig. 1 に示す。

テキストマイニングによる分析を行いやすくするために、地震発生 (2016 年 4

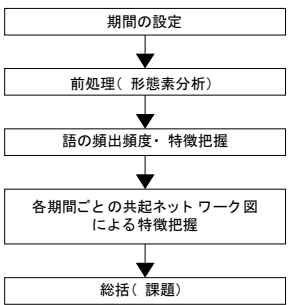


図 1 分析のフローチャート

表 1 対象期間および対象記事⁴⁾

	期間	記事数	内容 (例)
A 期	2016 年 4 月 14 日～ 2016 年 4 月 20 日	1588	気象庁によると、14 日夜、熊本県で震度 7 の地震があった。
B 期	2016 年 4 月 21 日～ 2016 年 5 月 4 日	1123	九州高校体育連盟は 28 日までに、6 月中旬から始まる九州高校大会で熊本県内で開催予定だった自転車、ラグビー、ソフトテニス、フェンシング、飛び込みを他県で開催することにした。地震の被害が大きく、会場も続いているため。日程の変更はなく、今後、代替地を決める。
C 期	2016 年 5 月 5 日～ 2016 年 5 月 18 日	579	13 日午前 1 時 3 分ごろ地震がありました。気象庁によると、震源地は熊本県熊本地方北緯 32.7 度、東経 130.7 度で、震源の深さは約 10 キロ。地震の規模はマグニチュード (M) 4.0 と推定される。震度 3 以上の地域震度 4＝熊本県熊本震度 3 以上の市町村震度 4＝熊本西区、宇城市▽震度 3＝熊本南区、八代市、宇土市、美里町、嘉島町、益城町、山都町、氷川町この地震による津波の心配はありません。
D 期	2016 年 5 月 19 日～ 2016 年 6 月 1 日	276	1 日午前 3 時 30 分ごろ地震がありました。気象庁によると、震源地は熊本県阿蘇地方北緯 33.0 度、東経 131.2 度で、震源の深さは約 10 キロ。地震の規模はマグニチュード (M) 2.8 と推定される。各地の震度は次の通り。震度 1＝菊池旭志、阿蘇一の宮、阿蘇波野、阿蘇、小国、産山、高森、阿蘇山 (熊本) 竹田 (大分) この地震による津波の心配はありません。
E 期	2016 年 6 月 2 日～ 2016 年 6 月 15 日	314	県は 10 日、熊本・大分地震による県内の建物被害が 3600 棟を超えたと発表した。各自治体は被災者から罹災 (りさい) 証明書の申請を受けて被害状況の調査を続けており、今後の調査でさらに増える可能性がある。県によると、被災した建物は 3603 棟になった。このうち住家被害は 3229 棟。全壊は別府市、宇佐市、九重町で各 1 棟。半壊は 106 棟で、由布市 61 棟、別府市 38 棟―など。瓦が落ちるなどの一部破損は別府、由布市を中心に 3120 棟に上った。倉庫など非住家の被害は 374 棟。全壊 17 棟、半壊 5 棟、一部破損が 352 棟となっている。

月 14 日) から 2016 年 6 月 15 日までの 2 ヶ月間を 5 つに分割し、それぞれ「A 期」, 「B 期」, 「C 期」, 「D 期」, 「E 期」とした。地震発生から 1 週間を A 期, 以後 2 週間ごとに, B 期, C 期, D 期, E 期と設定した。また, テキストマイニングによる分析を A 期～E 期それぞれに対して行う。

表 2 は, 対象期間内の地震発生回数^{5) 6)}を示している。地震発生から 8 週目にあたる E 期でも震度 5 弱の地震が発生し, 断続的な余震が 2 カ月以上続いていることがわかる。

3. テキストマイニングによる分析

3. 1 前処理

出現頻度やネットワーク分析等を行うにあたって, 語の出現頻度が僅かなものまでを対象とすると, 分析結果の解釈が困難になることから, 出現頻度の閾値を設定するのが一般的である。小林ら²⁾は閾値を「5」とすることで, 分析の解釈が十分可能であるとしている。しかし, 本研究では閾値の設定を行っていない。これは今後の課題とする。一方, 今回の分析対象

は新聞記事であるため誤字脱字は少ないと考えられる。よって, アンケート調査などの自由記述を対象として分析を行う際と同義語の変換処理や誤字等の削除・訂正については行っていない。

3. 2 語の出現頻度と特徴

表 3 には A 期～E 期における抽出語と出現頻度上位

30 位を示している。語の出現頻度は「震度」「地震」「熊本」「マグニチュード」などが多く, 上位 3 位は「震度」「地震」「熊本」が占めている。これは, 断続的に続く余震に関する記事が多いことが考えられる。また, A 期では地名や日時についての語が多く, B 期以降では「被災者」「公民館」「支援」「観光」「災害」などの語の出現頻度が高くなっており, 地震による被害や避難所または支援物資についての記事が増加していると考えられる。

4. 対象期間ごとの共起ネットワーク

対象期間における出現パターンの似通った語, すなわち共起の程度が強い語を線で結んだものを共起ネットワーク図という。また, 今回は Jaccard^{注2)}係数で測定した共起の程度に応じて, 共起関係をあらわす線の太さに変化し, 語の出現数に応じて, それぞれの語をあらわす円のサイズが変化すると設定した。新聞記事は毎日掲載記事が変わり, 地震発生時と 2 か月後の記事は, 掲載された情報や記事数が変化していることが予想される。そこで, 本研究では時間経過による共起関係の変化を見るため, 見出しとして「日付」を設定した。

テキストマイニングは文章全体を分析するため, 抽出語の取捨選択を行う必要がある。しかし, 共起ネットワーク図において「抽出語の取捨選択」や「描画する語の数」の閾値などが設定しにくいことが課題である。本研究では A 期～E 期を通して, 今回は「出現数による語の取捨選択」の閾値を 50 語, 「描画する共起関係」の閾値を 200 と設定した。分析対象の語選定基準は, KH Coder 内の初期設定から「名詞」「サ変名詞」「形容動詞」「固有名詞」「組織名」「人名」「地名」「ナイ形容」「副詞可能」「未知語」「感動詞」「動詞」「形容詞」「副詞」「名詞 C」を対象とした。

(1) A 期における共起ネットワーク図

図 2 は A 期 (2016 年 4 月 14 日～2016 年 4 月 20 日) における共起ネットワーク図である。中心において「地震」「震度」「熊本県」「マグニチュード」などが日付と強い共

表 2 期間内地震発生回数^{5) 6)}

月日	最大震度別回数							震度 1 以上を 観測した回数	累計
	1	2	3	4	5弱	5強	6弱		
4月14日	12	10	6	9	1			1	40
4月15日	30	51	19	10	1			1	112
4月16日	20	70	67	36	4	1	2	1	202
4月17日	29	70	28	11					138
4月18日	19	33	22	4	1				79
4月19日	24	33	20	2	1	1			81
4月20日	18	40	15	1					74
4月21日	9	36	11	2					48
4月22日	12	24	4	1					41
4月23日	14	13	1						28
4月24日	15	8	7						30
4月25日	9	14	1						24
4月26日	19	6	5						30
4月27日	34	11	4						49
4月28日	34	13	2	3					52
4月29日	15	10	7			1			33
4月30日	19	11	2						32
5月1日	24	9	3						36
5月2日	22	10	1						33
5月3日	18	5	2						25
5月4日	13	10	3	3					29
5月5日	23	10	3	3					39
5月6日	14	7	4						25
5月7日	13	6	2						21
5月8日	23	8	1						32
5月9日	20	3	3						26
5月10日	6	4	3						15
5月11日	12	3	1						16
5月12日	9	5	1	1					16
5月13日	12	4	1	1					18
5月14日	15	4	3						22
5月15日	12	1	2						15
5月16日	10	2							12
5月17日	6	3	1						10
5月18日	8	4	1						13
5月19日	5	3							10
5月20日	7	5	1						13
5月21日	8	3	3						14
5月22日	9	2	1						12
5月23日	7	1	1						9
5月24日	7	1							8
5月25日	7	2							9
5月26日	5	1							6
5月27日	6	2							8
5月28日	8	5							13
5月29日	2	1	1						4
5月30日	5	3							8
5月31日	2	1							3
6月1日	4	2							6
6月2日	8	5	3						16
6月3日	7	2	1						10
6月4日	2	1	1						4
6月5日	5	1	2						8
6月6日	8	2	1						11
6月7日	7	4	1						12
6月8日	10	4							14
6月9日	5	1							6
6月10日	5	2							7
6月11日	3	2							5
6月12日	7	2	1		1				11
6月13日	8	2	1						11
6月14日	3								3
6月15日	5	3							8
2ヶ月合計	757	607	273	89	8	4	3	2	1745

※空欄は「0」

表 3 抽出語と出現頻度

順位	A期		B期		C期		D期		E期	
	抽出語	出現頻度	抽出語	出現頻度	抽出語	出現頻度	抽出語	出現頻度	抽出語	出現頻度
1	震度	3445	地震	1603	震度	1034	地震	400	震度	441
2	地震	1869	震度	1512	地震	925	震度	399	地震	441
3	熊本	1329	熊本	922	熊本	629	熊本	266	熊本	308
4	熊本県	1009	熊本県	633	マグニチュード	370	規模	160	月	185
5	マグニチュード	873	マグニチュード	505	深い	367	マグニチュード	154	次	171
6	震源地	839	震源地	490	震源地	362	深い	154	規模	169
7	東経	830	気象庁	480	分ごろ地震	361	震源地	152	マグニチュード	163
8	分ごろ地震	791	東経	478	次	359	次	151	気象庁	163
9	気象庁	674	分ごろ地震	478	次	352	東経	151	震源地	163
10	日午前	669	深い	456	規模	351	分ごろ地震	150	深い	162
11	震源	628	規模	455	震源	348	震源	147	分ごろ地震	161
12	深い	625	心配	448	気象庁	347	津波	142	東経	159
13	大分	592	震源	440	心配	337	心配	142	震源	152
14	規模	597	次	433	津波	333	気象庁	143	心配	151
15	推定	572	津波	431	推定	332	津波	143	津波	151
16	心配	570	推定	421	各地	309	各地	140	推定	149
17	日午後	570	日午後	392	熊本県	304	支援	115	各地	142
18	地域	561	各地	375	日午後	236	月	114	熊本県	139
19	津波	560	大分	346	熊本県熊本地方北緯	228	被害	110	支援	123
20	次	553	日午前	323	日午前	197	熊本県熊本地方北緯	109	熊本県熊本地方北緯	110
21	熊本県熊本	541	被害	305	月	187	熊本県	107	日午後	100
22	八代	539	避難所	298	支援	141	県	106	日午前	98
23	北緯	465	支援	281	大分	136	日午前	89	大分	96
24	阿蘇	444	熊本県熊本地方北緯	261	地方	127	日午後	83	被害	93
25	福岡	411	月	246	阿蘇	126	熊本・大分地震	62	観光	84
26	地方	409	県	234	被害	125	災害	61	熊本・大分地震	79
27	宮崎	376	被災者	234	熊本	123	観光	58	八代	78
28	熊本県熊本地方北緯	366	公民館	209	地域	119	方門	58	阿蘇	77
29	長崎	337	北緯	204	北緯	119	市	56	県	75
30	上天草	329	地域	203	災害	107	地域	55	地域	70

起関係にあり、熊本地震や余震についての情報が多く、それぞれの日付とも共起関係が強いことが分かる。また、「大分」「大分両県」「大分県中部」など大分県に関する語が16日以降で共起関係を形成している。これは、一連の地震で、大分県中部を震源とした地震が発生していることに起因すると考えられる。20日以降は「支援」「対応」「必要」「影響」などの語が出現することから、時間経過とともに地震による影響についての記事が増えていると考えられる。

（２）Ｂ期における共起ネットワーク図

図3はB期（2016年4月21日～2016年5月4日）における共起ネットワーク図である。

A期と同様に「地震」「震度」「熊本県」などが日付と強い共起関係を示している。また、大分県内の地名も共起関係にあることから、大分県内における地震の影響も伝えて

いることが考えられる。また、「被災者」「避難所」「自宅」「被害」などが共起関係を形成しており、被災者の動向や現状を伝えていることが考えられる。

（３）Ｃ期における共起ネットワーク図

図4はC期（2016年5月5日～2016年5月18日）における共起ネットワーク図である。

「被災」「生活」「人」「話す」などが17日を中心に共起関係にあることから、依然として被災者の生活や影響が記事に記載されていることが考えられる。「支援」と「影響」が共起関係にあり、余震などにより被災地支援に影響が出ていることなども考えられる。

（４）Ｄ期における共起ネットワーク図

図5はD期（2016年5月19日～2016年6月1日）における共起ネットワーク図である。

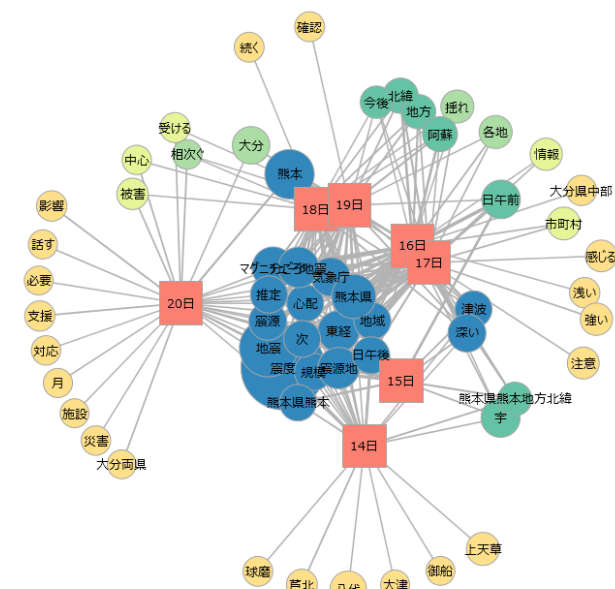


図2 A期の共起ネットワーク図

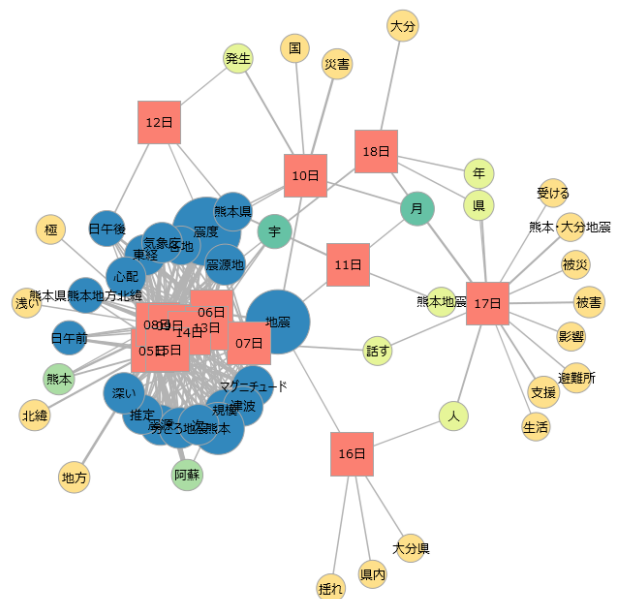


図3 B期の共起ネットワーク図

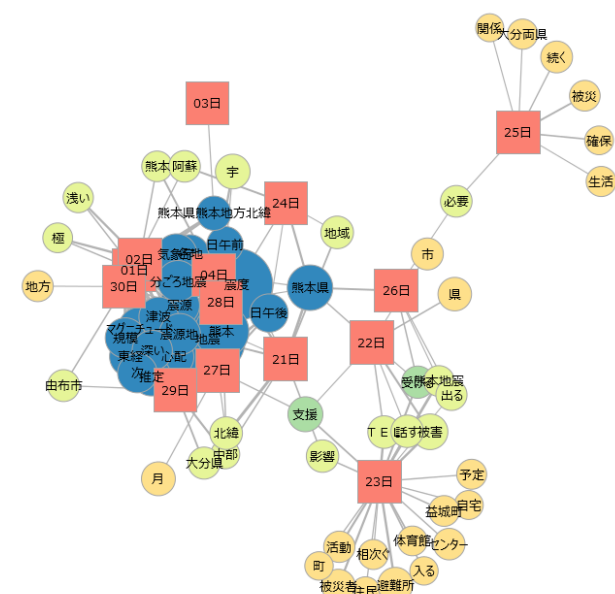


図4 C期の共起ネットワーク図

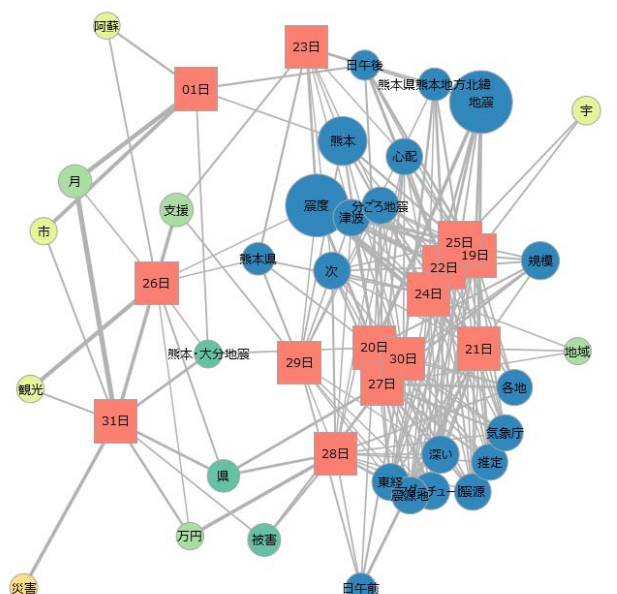


図5 D期の共起ネットワーク図

表1からもわかるように、時間経過に伴って地震関連の記事が少なくなること、語の出現数が減少している。しかし、A期と同様に「地震」「震度」「熊本県」などが強い共起関係を示していることから、余震に関する情報が地震関連の記事で大きな割合を占めていると考えられる。「観光」「万円」「被害」といった語の出現により、地震の風評被害による損失が記事の中に出現していると考えられる。

(5) E期における共起ネットワーク図

図6はE期(2016年6月2日～2016年6月15日)における共起ネットワーク図である。

A期と同様に「地震」「震度」「熊本県」などが強い共起関係を示している。一方で、「復興」「復旧」という語が出現し、共起関係を示している。

5. 総括

以上のように、本論では地方紙HPに掲載された熊本地震に関連した新聞記事に対して、A期～E期それぞれにおいてテキストマイニングによる分析を試みた。本研究による成果、分析を行う上での課題を以下にまとめる。

5. 1 本研究による成果

共起ネットワーク図より新聞記事における出現語の共起関係が明らかになった。地震直後は、「地震」「震度」などの情報が多くを占めるが、時間経過に伴って「避難所」「生活」「復旧」などの生活や影響をあらわす語が多く出現する。

5. 2 課題

本論では、新聞記事を対象にテキストマイニング分析を行い、共起ネットワーク図を作成した。分析の課題として次のことがあげられる。

(1) 出現頻度による閾値の設定

本研究では語の出現頻度による閾値の設定を行わず、出

現頻度が僅かな語も分析対象としている。よって、語の関係性などの解釈が困難な点や対象語数が膨大になることで、分析の有意性が下がる点が考えられる。

(2) 分析対象の語選定

新聞記事の一文一文は様々な品詞によって形成されている。しかし、分析対象の語選定を行わなければ、それ一語では意味を持たない語も分析対象に含まれてしまい、共起関係がうまく解釈できないことが考えられる。

(3) 共起ネットワーク図作成時の設定

A期～E期までの出現語数の違いによって、共起ネットワーク図は変化することが考えられる。

(4) 共起コミュニティの形成

福井ら⁷⁾は、サブグラフ検出による分類を行うことで共起ネットワークの中でも、比較的強くお互いに結びついてあるグループを「共起コミュニティ」としている。本研究において「共起コミュニティ」について分析を行うことができれば、新聞記事の傾向を把握できると考える。

このような課題を踏まえることで、メディアがどのような情報をいつ伝えたのか、また、住民が求める情報との乖離、さらに地震後の情報の整理が行えることができると考えられる。

【補注】

- 注1) KH coderは樋口耕一氏(立命館大学)により開発されたテキストマイニングを行うためのフリーソフトウェアである。
注2) Jaccard 係数とは、語の共起の強さを図る指標

【参考文献】

- 平成28年(2016年)熊本県熊本地方を震源とする地震に係る被害状況等について：<http://www.bousai.go.jp/updates/h280414jishin/index.html> (2016.9.24 最終閲覧)
- 小林祐司, 寺田充伸, 佐藤誠治：テキストマイニングを活用したアンケートにおける自由回答の分析と生活環境評価, 日本建築学会計画系論文集, 第77巻, 第671号, pp.85-93, 2012.1
- 伊藤香織, 青野貞康, 大森宣暁：首都圏における震災時帰宅立ち寄り行動の実証研究—東日本大震災に関する研究 web アンケート調査に基づく分析—, 日本都市計画学会都市計画論文集, Vol.48, No.3, 2013.10
- 大分合同新聞 HP：<http://www.oita-press.co.jp/>
- 気象庁：「平成28年(2016年)熊本地震」について(第40報)—平成28年6月12日2時08分頃の熊本県熊本地方の地震—<http://www.jma.go.jp/jma/press/1606/13a/kaisetsu201606130015.pdf> (2016.9.24 最終閲覧)
- 気象庁：震度データベース検索(熊本県と大分県発生データの)<http://www.data.jma.go.jp/svd/eqdb/data/shindo/index.php> (2016.9.24 最終閲覧)
- 福井美弥, 阿部浩和, 橋寺知子：産業遺産施設の保存活用の現状と事業主体の役割—大阪・兵庫の繊維産業遺産施設8施設を対象として—, 日本建築学会計画系論文集, 第78巻, 第687号, pp.1067-1076, 2013.5

- *1 大分大学院工学研究科博士前期課程工学専攻 大学院生
*2 大分大学工学部福祉環境工学科・建築コース 准教授・博士(工学)

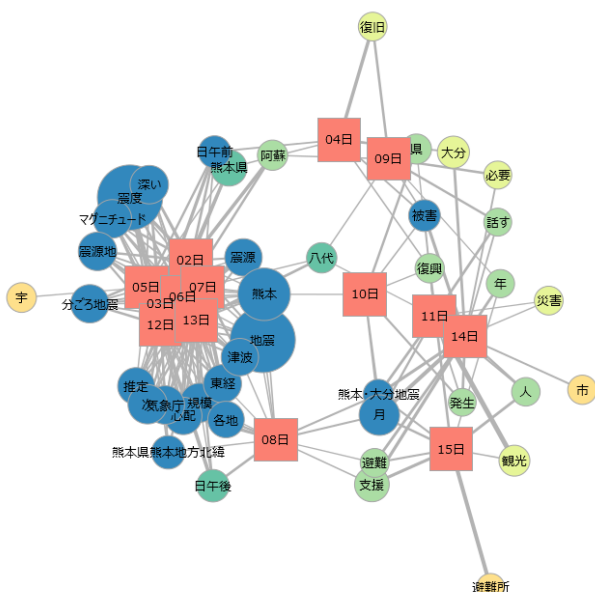


図6 D期の共起ネットワーク図