

NCR2018における規定間参照関係：RDF データに対するネットワーク分析の適用

谷口 祥一（慶應義塾大学文学部）
taniguchi@z2.keio.jp

橋詰 秋子（実践女子大学短期大学部）
hashizume-akiko@jissen.ac.jp

RDF データ化した NCR2018 を活用し展開するという観点から、3つの章の個別規定間における参照関係に着目し、ネットワーク分析を試行した。グラフ（ネットワーク）を構成するノードは、条項番号単位、条項番号の階層上の3桁レベル単位、個々の条項が対象とする NCR2018 エレメント単位の3種類を設定した。ノード設定単位ごとに参照関係のグラフを生成し、①グラフの基本的特徴量の確認、②中心性の算出、③グラフ分割の実施という3点から分析を行い、ネットワーク分析の有効性を検証した。

1. はじめに

発表者はこれまでの研究発表において、NCR2018 の記述規則を RDF によるデータ表現とするために必要な検討を行った¹⁾²⁾。JLA 目録委員会が公開した NCR2018 語彙の RDF 定義（エレメントや関連指示子を示す RDF プロパティなど）にそのまま接続できることを前提とした上で、①個別規定を RDF クラスとする場合の設定単位、②規定間の参照関係の表現、③別法の表現、④例示の記録法などを検討した。加えて、⑤個別規定から当該規定に従って作成されたメタデータへの参照、⑥逆にメタデータから実際に適用された個別規定への参照、⑦国立国会図書館による適用細則の RDF データ化および対応する NCR 規定との接続などについて、それぞれ必要な検討項目を示し、可能な選択肢を提示した。

本発表は、RDF データ化した NCR2018 を活用し展開するという観点から、個別規定間の参照関係に着目したネットワーク分析を試行する。これによって、メタデータ記述規則に対するネットワーク分析の適用可能性および有効性を検証するとともに、入念かつ膨大に設定されている NCR2018 内の参照関係の実態やその特徴を明らかにする。本研究では、NCR2018 の第2章（表現形の属性の記録）、第4章（著作の属性の記録）、および第6章（個人の属性の記録）を対象とした。

2. 個別規定間の参照指示の抽出と表現

発表者のこれまでの研究成果を踏まえて、NCR2018 の各章のテキストファイルを個別条項単位（場合によっては、それをさらに分割した単位）の RDF トリプルデータに機械的に変換した。この変換データに対して、人手による確認と誤変換等の修正を行った。

本研究で着目する参照指示は、プロパティ「referredInstrct」を用いて表現しており、1

つの条項例を下記に示す。

#2.1.1.2.3 ルビ

:Ncr2.1.1.2.3

rdf:type :NcrInstruction ;

:instrctNo "#2.1.1.2.3";

rdfs:label "タイトル - 本タイトル - 記録の方法 - ルビ";

:instrctFor

<<http://jla.or.jp/term/ncr2018/E200002>>;

:instruction "情報源に表示されたタイトルに付記されたルビは、本タイトルに含めない。識別またはアクセスに重要な場合は、(以下、省略)";

:referredInstrctStmnt " (参照: #2.1.9.1.1h)、#2.1.9.2.1 を見よ。) ";

:referredInstrct :Ncr2.1.9.1.1h, :Ncr2.1.9.2.1 ;

:instrctAppExample "青い思想 (情報源の表示: 青い思想(こころ)) ";

:upper :Ncr2.1.1.2 .

NCR2018 の参照指示には、①独立した参照指示（例：（参照: #2.1.1.4b）、#2.1.1.4 別法 b）、#2.2.0.6 を見よ。)) と、②規定文内における参照指示（例：「原資料のタイトルが同一の情報源に表示されている場合は、#2.1.1.3 に従う。」）とがあり、本研究では両者を RDF トリプルの抽出対象とした。参照指示は、参照先の個々の条項番号を記したものと範囲を示したもの（例：「#2.10.1.2~#2.10.1.2.4 別法」）があり、後者については当該範囲に含まれる規定のうち、実際の参照先条項と見なすべきものの解釈に幅がありうる。そこで、実際の参照先条項と見なすべきものに関して以下の方針を定め、それに依拠した抽出を行った。

①範囲指定における先頭の条項番号と末尾の条項番号とが階層レベルで一致しており、かつ下記②に該当しない場合には、その中間に該当する条項について同じ階層レベルのもののみ採用する。また、それらの下位レベルの条項は参照しない。例えば、「#4.3~#4.7」の場合には、#4.3, #4.4, #4.5, #4.6, #4.7 の参照とする。

②先頭条項番号と末尾条項番号とが階層レベルで一致しているが、一部の中間レベルが異なる

る条項番号の場合には、共通部分からなるパターンを設定し、それに該当する条項のみ参照する。例えば、「#4.16.0.1.2～#4.20.0.1.2」は、共通パターン「#4.?.0.1.2」とし、#4.16.0.1.2、#4.17.0.1.2、#4.18.0.1.2、#4.19.0.1.2、#4.20.0.1.2のみ参照する。

③先頭条項番号と末尾条項番号とが階層レベルで一致しない場合には、a)先頭番号が階層上位、末尾番号が階層下位のときには、先頭番号と同列の条項を順次参照し、加えて末尾番号については指定された下位レベルの条項まで順次参照する。例えば、「#1.11～#1.12.3」は、#1.11、#1.12、#1.12.1、#1.12.1 別法、#1.12.2、#1.12.2 別法、#1.12.3 への参照とする。b)先頭番号が階層下位、末尾番号が階層上位のときには（例：「#4.8.3～#4.12」）、先の処理 a と逆にする。c)これら以外のパターンは解釈不能であり、参照として採用していない。

3. 参照関係が構成するグラフの特徴量

RDF トリプルとして表現された参照関係をグラフ（ネットワーク）と捉えると、①参照元の規定から参照先規定をもつ有向グラフ、かつノード（頂点）の設定単位によっては、②同一の参照関係が発生し、重み付きグラフ（または多重グラフ）、そして③参照元と参照先が同一となる自己ループありグラフとなる。参照先ノードには他の章の規定も含まれる。

参照元と参照先のノードの設定単位として、①条項番号単位（例：「:Ncr2.2.0.4.2」）、②条項番号の階層上の 3 桁レベル単位（例：「:Ncr2.2.0」）、③個々の条項が対象とする NCR2018 エレメント単位（例：「責任表示」）を採用した。①の場合、「:Ncr2.1.9.1.1h」は「:Ncr2.1.9.1.1」に統合され、②の場合には「:Ncr2.1.9」となる。さらに、②においては、条項番号末尾の A、B などは削除し、「別法」・「任意省略」・「任意追加」は本則に統合した。また、③は個々の条項が特定のエレメントを対

象としていない「通則」の場合には、エレメントに代えて、「体现形 - 通則 - 情報源」などを適宜設定した。各章を構成する条項数と、参照指示をもつ条項数、もたない条項数、抽出したトリプル数（参照指示回数）などを第 1 表に示した。併せて、条項番号 3 桁レベル単位にまとめ上げた場合の参照関係の集計結果を、第 1 表下側に示した。

対象とした 3 つの章における参照関係から形成したグラフのノード数とエッジ数（辺の数；参照関係の種類数）、エッジの出現回数（参照関係の延べ出現回数）を、第 2 表に示した。この結果から、章による揺れがあるとはいえ、対象としたいずれの章においても、ほとんどの参照指示は当該章内に向けたものであり、他の章の規定への参照は条項番号単位で 20%未満であることが明らかとなった。

これらの参照関係をグラフの特徴という観点から見ると、特に第 2 章の規定群で条項番号単位においては、一部のノードが他の多数のノードとエッジで連結しており、他の大部分は少数のノードとしかつながっていない、すなわち次数分布が「べき法則」に従っている（両対数グラフではほぼ直線になる）、スケールフリーグラフと見なすことができる（クラスタ係数は 0.045 と小さい）。他方、条項番号 3 桁単位やエレメント単位有的时候には、クラスタ係数が比較的大きく（0.207 と 0.237）、小さい平均パス長（1.652 と 1.903）をもつ、スモールワールドグラフに近いと判断できる。第 4 章・第 6 章の規定群については、いずれのノード設定単位においても、おおむねスモールワールドグラフと捉えられる。グラフの特徴量である密度（density）、推移性（transitivity）、相互性（reciprocity）、クラスタ係数（local clustering coefficient）、平均パス長（average shortest path length）を第 3 表にまとめた。なお、平均パス長については、非連結であり計算できない場合もあった。

第 1 表 参照指示の抽出結果

	全条項数	参照指示あり条項数	参照指示なし条項数	抽出トリプル数	条項当たり抽出トリプル数		
					最大値	平均値	SD
条項番号単位							
第 2 章	1254	594 (47.4%)	660 (52.6%)	2137	42	1.70	3.92
第 4 章	277	140 (50.5%)	137 (49.5%)	807	49	2.91	6.14
第 6 章	148	59 (39.9%)	89 (60.1%)	252	37	1.70	4.79
条項番号 3 桁単位							
第 2 章	218	130 (59.6%)	88 (40.4%)	2137	130	9.80	18.34
第 4 章	89	48 (53.9%)	41 (46.1%)	807	146	9.07	23.90
第 6 章	102	42 (41.2%)	60 (58.8%)	252	42	2.47	6.53

4. 参照関係が構成するグラフの中心性

NCR2018 各章には参照関係の核となる条項やエレメントが存在するという仮説を立て、この仮説を検証するために、3つの章の参照関係グラフに対して各種の中心性尺度を算出した。具体的には、次数中心性 (degree centrality)、入次数中心性、出次数中心性、固有ベクトル中心性、Katz 中心性、PageRank、媒介 (betweenness) 中心性、近接 (closeness) 中心性の各尺度を算出した。これらの尺度はそれぞれ異なる側面から中心性を計算しているが、いずれの章においても各尺度によって上位となったノード群には重なる部分がある程度見られた。これらの上位のノード群は各章の参照関係ネットワークの核に位置付けられるものと考えられる。

さらに、中心的なノード群の傾向を探るために、①個々の中心性尺度の値が上位 10 位までのノード (10 位のものが複数あれば、それらすべて) を採用し、②採用されたノード群を対象にクラスタリングを適用し、各種中心性尺度値において類似のパターンをとるものを同定した。距離測定には各尺度値に対してユークリッド平方距離を適用し、Ward 法と群平均法 (グループ間平均連結法) により階層的クラスタリングを実施した。

例えば、第 2 章の条項番号単位のグラフにおいては、上記の中心性尺度のいずれかが上位 10 位までのノード群は 68 の条項となり、重なりはそれほど大きくない。これらは体現形全体の通則や個々のエレメントの通則などが多い。それ以外では、体現形の注記、キャリアに関する注記、数量、転記 (第 1 章 #1.10)、体現形間の関連 (第 43 章 #43.3) などが含まれていた。これらノード群を Ward 法によりクラスタリングすると、最も下位レベルで 4 つのクラスタに分かれた (クラスタ 1: 条項数 36、クラスタ 2: 19 条項、クラスタ 3: 3 条項、クラスタ 4: 10 条項)。クラスタ 1 は体現形の情報源、タイトルの誤表示など、記録の方法に関する多様な規定群が含まれている一方、クラスタ 2 は転記の条項等、クラスタ 3 は数量の通則、クラスタ 4 は「テキストの数量」の下位規定となった。これらのクラスタから、記録の方法、転記、数量、テキストの数量に関わる条項が 2 章の参照関係の中心となる傾向が示唆された。

他方、第 4 章の条項番号単位の場合には、各種中心性尺度上位 10 位のノード群は合計 30 に限定され、重なりが大きく、それらに対する Ward 法・群平均法のクラスタリング結果はい

ずれも下記の 5 クラスタとなった。

クラスタ 1 (9 規定): 著作の年代に基づく優先タイトルの選択と言語に依拠した記録の方法

(別法)、体現形の優先情報源に関する通則

クラスタ 2 (9 規定): 著作の集合と部分における優先タイトルの選択と記録の方法、音楽作品の優先タイトルの選択と記録の方法

クラスタ 3 (9 規定): 著作の記録の目的と方法の通則、音楽作品と法令等の通則

クラスタ 4 (2 規定): 著作の年代に基づく優先タイトルの選択と言語に依拠した記録の方法

クラスタ 5 (1 規定): 文字種に依拠した著作の優先タイトルの選択

上記の結果から、第 4 章の参照関係の中心は優先タイトルの選択に関わる条項群にあることが示唆された。同様に、他の章や他のノードレベル設定において、クラスタの構成が比較的解釈しやすい構図で示された場合もある。

5. 参照関係が構成するグラフの分割

参照関係の密度等に基づき各章の条項をいくつかのまとまりに分けることができるのではないかと仮定し、参照関係グラフの分割を試行した。与えられた分割数や分割後のサイズに基づいてグラフを分割することをネットワーク分割、ネットワークの密な部分を (分割数などを指定せずに) 抽出することをコミュニティ抽出と呼ぶ。これらに該当する複数手法を試行した結果、分割数を指定するネットワーク分割の結果が本研究にとって適切と判断した。有向グラフに適用可能な Girvan-Newman アルゴリズム、無向グラフとした上で適用可能な非同期流体 (Asynchronous Fluid) コミュニティ・アルゴリズムを採用した。分割数を 10, 15, 20, 25 としてその結果を確認した。前者のアルゴリズムはサブグラフを順次分割していくのに対して、後者は分割数ごとにサブグラフをその都度構成するため、サブグラフを構成するノードの入れ替えが発生する。

条項番号単位、条項番号 3 桁単位、エレメント単位において試行したところ、いずれの章の規定群についても、非同期流体アルゴリズムによる結果が解釈しやすいものとなった。例えば、第 2 章の条項番号 3 桁単位の規定群を 20 グループに分割した場合には、第 4 表のような結果となった。

以上の結果から、記述規則における規定間の参照関係にネットワーク分析が問題なく適用でき、かつ部分的にせよ、有効な結果が得られることが確認できた。

引用文献

1) 谷口祥一. NCR2018 と RDA の記述規則の RDF データ化. 第 68 回日本図書館情報学会研究大会発表論文集. 2020, p.41-44.

2) 谷口祥一, 橋詰秋子. NCR2018 の RDF データ化: 記述規則とメタデータの接続等による展開. 2021 年度日本図書館情報学会春季研究集会発表論文集. 2021, p.25-28.

第 2 表 参照関係が形成するグラフの概要

条項番号単位	第 2 章		第 4 章		第 6 章	
	ノード数	エッジ出現回数	ノード数	エッジ出現回数	ノード数	エッジ出現回数
すべての参照	2064	2137	750	807	251	252
当該章内への参照	1781 (86.3%)	1851 (86.6%)	626 (83.5%)	683 (84.6%)	211 (84.1%)	212 (84.1%)
同一条項参照	0 (0.0%)	0 (0.0%)	1 (0.1%)	2 (0.2%)	0 (0.0%)	0 (0.0%)
上位条項参照	5 (0.2%)	13 (0.6%)	20 (2.7%)	24 (3.0%)	2 (0.8%)	2 (0.8%)
下位条項参照	183 (8.9%)	198 (9.3%)	90 (12.0%)	113 (14.0%)	9 (3.6%)	9 (3.6%)
前方条項参照	992 (48.1%)	1011 (47.3%)	271 (36.1%)	285 (35.3%)	111 (44.2%)	112 (44.4%)
後方条項参照	601 (29.1%)	629 (29.4%)	244 (32.5%)	259 (32.1%)	89 (35.5%)	89 (35.3%)
他章への参照	283 (13.7%)	286 (13.4%)	124 (16.5%)	124 (15.4%)	40 (15.9%)	40 (15.9%)
条項番号 3 桁単位	ノード数 225		ノード数 113		ノード数 100	
すべての参照	813	2137	307	807	138	252
当該章内への参照	634 (78.0%)	1851 (86.6%)	223 (72.6%)	683 (84.6%)	112 (81.2%)	212 (84.1%)
同一条項参照	30 (3.7%)	385 (18.0%)	8 (2.6%)	282 (34.9%)	3 (2.2%)	18 (7.1%)
前方条項参照	348 (42.8%)	933 (43.7%)	92 (30.0%)	174 (21.6%)	52 (37.7%)	105 (41.7%)
後方条項参照	256 (31.5%)	533 (24.9%)	123 (40.1%)	227 (28.1%)	57 (41.3%)	89 (35.3%)
他章への参照	179 (22.0%)	286 (13.4%)	84 (27.4%)	124 (15.4%)	26 (18.8%)	40 (15.9%)
エレメント単位	ノード数 187		ノード数 83		ノード数 49	
すべての参照	849	2137	322	807	139	252
同一エレメントへの参照	31 (3.7%)	480 (22.5%)	8 (2.5%)	253 (31.4%)	5 (3.6%)	18 (7.1%)
異なるエレメントへの参照	818 (96.3%)	1657 (77.5%)	314 (97.5%)	554 (68.6%)	136 (97.8%)	234 (92.9%)

第 3 表 参照関係が形成するグラフの特徴量

	第 2 章			第 4 章			第 6 章		
	条項番号単位	条項番号 3 桁単位	エレメント単位	条項番号単位	条項番号 3 桁単位	エレメント単位	条項番号単位	条項番号 3 桁単位	エレメント単位
密度	.002	.016	.024	.010	.024	.047	.012	.014	.059
推移性	.022	.064	.079	.077	.040	.062	.047	.024	.049
相互性	.172	.362	.327	.104	.091	.081	.120	.116	.129
クラスタ係数	.045	.207	.237	.140	.182	.149	.108	.057	.152
平均パス長	—	1.652	1.903	—	.802	1.270	—	—	.673

第 4 表 ネットワーク分割結果例 (第 2 章、条項番号 3 桁単位)

#2.7.0~#2.7.5 (製作表示)、#2.41.7、#44.3.3	#2.15.0、#2.16.0~#2.32 (キャリア情報; 途中欠番あり)
#2.5.0~#2.5.5 (出版表示)、#2.8、#2.41.5/9	#2.29.0~#2.29.8 (録音の特性)、#2.33.0
#2.8.0~#2.8.5 (非刊行物の制作表示)、#2.41.8	#2.5.3 (出版者)、#2.5.4 (並列出版者)、#2.10、#44.3.1
#2.4.0~#2.4.8 (逐次刊行物の順序表示)、#2.13.3	#2.18.0~#2.28.0 (キャリア情報; 途中欠番あり)
#2.1.0~2.1.9 (タイトル)、#2.10.1~#2.10.4	#2.1 (タイトル)、#2.3 (版表示)、#2.9 (著作権日付)
#2.14~#2.33 (キャリア情報; 途中欠番あり)	#2.34.0~#2.34.2 (体現形の識別子)、#2.39 (URL)
#2.4.1~#2.4.6 (逐刊の順序表示)、#2.41.4	#2.11 (下位レベルの記録)、#43.3 (体現形間の関連)
#2.2.0 (責任表示)、#2.3.3~#2.3.8 (版表示)	#2.1.1 (本タイトル)、#2.2.1~#2.2.2、#2.10.6~#2.10.7
#2.6.0~#2.6.5 (頒布表示)、#2.41.6、#44.3.2	#2.2 (責任表示)、#2.17.0~#2.17.5 (数量)、#2.12
#2.0.3~#2.0.4 (体現形一通則)、#2.3.0~#2.3.4	#2.32.0~#2.32.7 (デジタル・ファイルの特性)、#2.29.5

※紙幅の都合から一部条項を省略