

# アテンションダイナミクスに基づいた料理レシピ群の協調構造の分析

藤原稜<sup>†,a</sup>

熊野雅仁<sup>‡,b</sup>

小堀聰<sup>‡,c</sup>

木村昌弘<sup>‡,d</sup>

† 龍谷大学大学院理工学研究科電子情報学専攻 ‡ 龍谷大学理工学部電子情報学科

a) *t18m020@mail.ryukoku.ac.jp* b) *kumano@rins.ryukoku.ac.jp*

c) *kobori@rins.ryukoku.ac.jp* d) *kimura@rins.ryukoku.ac.jp*

**概要** 近年、料理レシピ共有サイトから得られる大規模データを用いて、人々の日常生活における料理レシピを科学的に分析する研究が注目されている。我々は最近、ソーシャルメディアにおけるアイテム群のアテンションダイナミクスのモデル化に対して、アイテム群の協調構造を組み込んだ CHP(Cooperative Hawkes Process) モデルとその学習法を与え、料理レシピ共有サイトの実データを用いた実験で有効性を示した。本研究では、アテンションダイナミクスに基づいた料理レシピ群の協調構造の特徴を明らかにすることを目指し、CHP モデルにより抽出された協調グループに属するレシピについて、代表的な利用食材や食材ネットワーク構造の観点から分析する。

**キーワード** ソーシャルメディアマイニング、確率過程モデル、協調構造、食材ネットワーク分析

## 1 はじめに

ソーシャルメディアは、Web 空間における重要なコミュニケーションの場として発展し、人々の日常生活にも大きな影響を与えつつある。近年、ソーシャルメディアの一つとして料理レシピ共有サイトが登場し、日常の家庭料理レシピに関する大規模データが得られるようになり、利用食材などの観点から人々の日常生活における料理レシピを科学的に分析する研究が活発化している [1-5]。料理レシピ共有サイトにおいてユーザは、自分が家庭料理で創作した料理レシピを投稿できるだけでなく、他のユーザが投稿した自分好みの料理レシピに対し賛意メッセージを送り、それを共有することも可能である。そして、投稿された料理レシピは、他のユーザ達からの共有というアテンションを通してポピュラリティを獲得していくことになる。したがって、このようなオンラインアイテムのアテンションダイナミクスをモデル化する研究、すなわち、ソーシャルメディアにおいてアイテムが共有されポピュラリティを獲得していく過程をモデル化する研究が、近年、ソーシャルメディアマイニングの分野で注目されている [6-8]。

オンラインアイテム群のアテンションダイナミクスのモデルとして、我々は最近の論文 [9] でアイテム群の協調構造を組み込んだ CHP(Cooperative Hawkes Process) モデルを提案した。CHP モデルは、Hawkes 過程 [10] と Dirichlet 過程 [11] を新たなやり方で融合することにより構築され、各アイテムに対する共有イベントの時系列を生成するものである。我々は、共有イベント系列の観測データから CHP モデルを効率よく推定する手法を

開発するとともに、アイテム群の協調構造を抽出する手法および、各アイテムの近い将来のポピュラリティを予測する手法を与えた。そして、料理レシピ共有サイトの実データを用いた実験により、CHP モデルが予測性能においてアイテム間に相互作用のない Hawkes 過程モデル [10] および多変量 Hawkes 過程モデル [12] を上回ることを実証し、協調構造を考慮する CHP モデルの有効性を示した。また、CHP モデルにより共有イベント系列から料理レシピ群の協調構造を抽出し、協調グループ間の影響関係を分析した。

しかしながら、CHP モデルにより抽出された協調グループに属するレシピで利用されている食材の特徴については、これまで十分な分析がなされていなかった。本研究では、CHP モデルにより抽出された協調グループ群について、各グループに属するレシピの代表的な利用食材や食材ネットワーク構造 [2, 13] の観点から詳細に分析し、アテンションダイナミクスに基づいたレシピ群の協調構造の特徴を明らかにすることを目指す。

## 2 アテンションダイナミクスモデル

対象とする料理レシピ全体の集合を  $\mathcal{A}$  とする。期間  $[0, T)$  での  $\alpha \in \mathcal{A}$  に対する共有イベント系列の発生過程（アテンションダイナミクス）をモデル化することを考える。ここに、 $T$  はそれほど大きくない正数（例えば、2,3 カ月）である。

### 2.1 CHP モデルの定義

任意のレシピ  $\alpha \in \mathcal{A}$  に対して、時刻  $t$  までのその共有イベントを計数過程  $N_\alpha(t)$  としてモデル化する。ただし、 $N_\alpha(t)$  は期間  $[0, t)$  内での  $\alpha$  の共有イベント数を表す。任意の  $\alpha \in \mathcal{A}$  に対して、時刻  $t$  までのその共有

イベント系列を,

$$\mathcal{T}_\alpha(t) = \{(t_n, \alpha); n = 1, \dots, N_\alpha(t)\}$$

とし, また, 時刻  $t$  までの  $\mathcal{A}$  に対する共有イベント系列を,

$$\mathcal{T}(t) = \{(t_n, \alpha_n); n = 1, \dots, N(t)\}$$

とする.  $N(t) = \sum_{\alpha \in \mathcal{A}} N_\alpha(t)$  である.

CHP モデルは, 多変量 Hawkes 過程に  $\mathcal{A}$  内の協調構造  $Z = \{z(\alpha); \alpha \in \mathcal{A}\}$  を組み込むことにより構築される. ここに,  $\mathcal{A} = \bigcup_{k=1}^K \mathcal{A}_k$  (disjoint union) であり,  $z(\alpha') = k$  であるのは,  $\alpha' \in \mathcal{A}_k$  であるときに限る. CHP モデルの強度関数  $\lambda_\alpha(t)$  は,

$$\lambda_\alpha(t | Z) = \mu_\alpha + \sum_{(t_n, \alpha_n) \in \mathcal{T}(t)} w_{\alpha, z(\alpha_n)} \exp \left\{ -\gamma_{z(\alpha_n)}(t - t_n) \right\}$$

で与えられる. ここに,  $\mu_\alpha > 0$  は  $\alpha$  の魅力を表し,  $w_{\alpha, k} > 0$  は  $\mathcal{A}_k$  から  $\alpha$  への影響度を表し,  $\gamma_k > 0$  は  $\mathcal{A}_k$  からの影響度の時間減衰率を表している. すなわち, 協調グループ  $\mathcal{A}_k$  に属する各レシピは, 任意のレシピ  $\alpha$  に  $w_{\alpha, k}$  という形で等しく影響を及ぼし, さらに同一の時間減衰率  $\gamma_k$  をもっている.

一般に, 協調グループの数  $K$  を事前に決定することは困難であるため, 観測データから  $K$  の値を決定できることが望ましい. したがって, CHP モデルでは, Dirichlet 過程を用いて  $\mathcal{A}$  内の協調構造を多変量 Hawkes 過程にノンパラメトリックベイズ形式で組み込むことにより, 任意の  $\alpha \in \mathcal{A}$  の共有イベント系列を生成する (詳細は [9] を参照).

## 2.2 CHP モデルの学習

観測系列  $\mathcal{T}(T)$  から CHP モデルを学習する. Dirichlet 過程は Chinese restaurant 過程 (CRP) [11] と等価であることから, CRP に基づく近似推論アプローチを用いる. また, 計数過程に関する重ね合わせの原理を用いて学習アルゴリズムを簡易化する. そのために, 第  $n$  共有イベント  $(t_n, \alpha_n)$  が第  $x_n$  共有イベント  $(t_{x_n}, \alpha_{x_n})$  によって引き起こされたことを表す潜在変数の集合  $X = \{x_n; n = 1, \dots, N(T)\}$  を導入する. ここに,  $x_n = 0, 1, \dots, n-1$  であり,  $x_n = 0$  は第  $n$  共有イベント  $(t_n, \alpha_n)$  がアイテム  $\alpha_n$  の固有の魅力によって引き起こされたことを意味する. そして, ベイズ的手法を用いて  $\mathcal{T}(T)$  から, CHP モデルのパラメータ  $K$ ,  $\{\mu_\alpha\}$ ,  $\{\gamma_k\}$ ,  $\{w_{\alpha, k}\}$  の推定値  $K^*$ ,  $\{\mu_\alpha^*\}$ ,  $\{\gamma_k^*\}$ ,  $\{w_{\alpha, k}^*\}$  を求める (詳細は [9] を参照).

このとき, 協調構造  $Z$  の推定値として,  $\alpha \in \mathcal{A}$  と  $k = 1, \dots, K^*$  に対し事後確率,

$$\theta_{\alpha, k} = P(z(\alpha) = k | \mathcal{T}(T), \{\mu_\alpha^*\}, \{\gamma_k^*\}, \{w_{\alpha, k}^*\})$$

が得られる. また, 潜在変数の集合  $X$  の推定値として,  $n = 1, \dots, N(T)$ ,  $i = 0, 1, \dots, n-1$  に対し事後確率,

$$\xi_{n, i} = P(x_n = i | \mathcal{T}(T), \{\theta_{\alpha, k}\}, \{\mu_\alpha^*\}, \{\gamma_k^*\}, \{w_{\alpha, k}^*\})$$

が得られる.  $\theta_{\alpha, k}$  はレシピ  $\alpha$  の協調グループ  $\mathcal{A}_k$  への帰属度を表し,  $\xi_{n, i}$  は第  $i$  共有イベント  $(t_i, \alpha_i)$  から第  $n$  共有イベント  $(t_n, \alpha_n)$  への影響を測定していることに注意する.

## 3 協調グループの分析法

事後確率  $\{\theta_{\alpha, k}\}$  に基づいて, 各協調グループ  $\mathcal{A}_k$  に属するレシピで利用されている食材について分析する.

本研究では, まず, Teng ら [2] に従い, 食材ネットワークとして ingredient complement network を構築し, それを可視化分析するとともに, Rosvall と Bergstrom [13] のネットワーククラスタリング法を適用して詳細に分析する. なお, 食材ネットワークの可視化法としては, Jacomy ら [14] の Force Atlas2 を用いる. ここに, 食材ペア  $a, b$  に対して, その結合重み PMI( $a, b$ ) は,

$$\text{PMI}(a, b) = \log \frac{P(a, b)}{P(a)P(b)}$$

で定義され, ingredient complement network 構築では統計的に特異なリンクのみを残す. ただし,

$$\begin{aligned} P(a, b) &= \frac{\# \text{ of recipes containing } a \text{ and } b}{\# \text{ of recipes}} \\ P(a) &= \frac{\# \text{ of recipes containing } a}{\# \text{ of recipes}} \\ P(b) &= \frac{\# \text{ of recipes containing } b}{\# \text{ of recipes}} \end{aligned}$$

であり, PMI は, 2 つの食材が独立にレシピに現れる確率と比較して, それらが同じレシピに現れる確率がどのくらい大きいかを測定している. 食材分析では, 構築した ingredient complement network における食材の次数に着目して分析を行う. 食材ネットワークのクラスタ分析では, 本研究では, 協調グループ  $\mathcal{A}_k$  に関して PMI を計算する際, 事後確率  $\{\theta_{\alpha, k}\}$  に基づいて  $\mathcal{A}_k$  に属するレシピやそこで利用されている食材を取り扱う. すなわち,  $\mathcal{A}_k$  という観点では, 各  $\alpha \in \mathcal{U}$  をレシピ 1 個分として扱うかわりにレシピ  $\theta_{\alpha, k}$  個分として扱う.

## 4 実験

### 4.1 実験設定

実験データとして, 料理レシピ共有サイト Cookpad の実データを用いた. Cookpad においてユーザは創作した料理レシピを投稿でき, また, 別のユーザはそれらの料理レシピに賛意を表すメッセージ「つくれぽ」を投稿できる. 料理レシピとつくれぽをそれぞれアイテムと

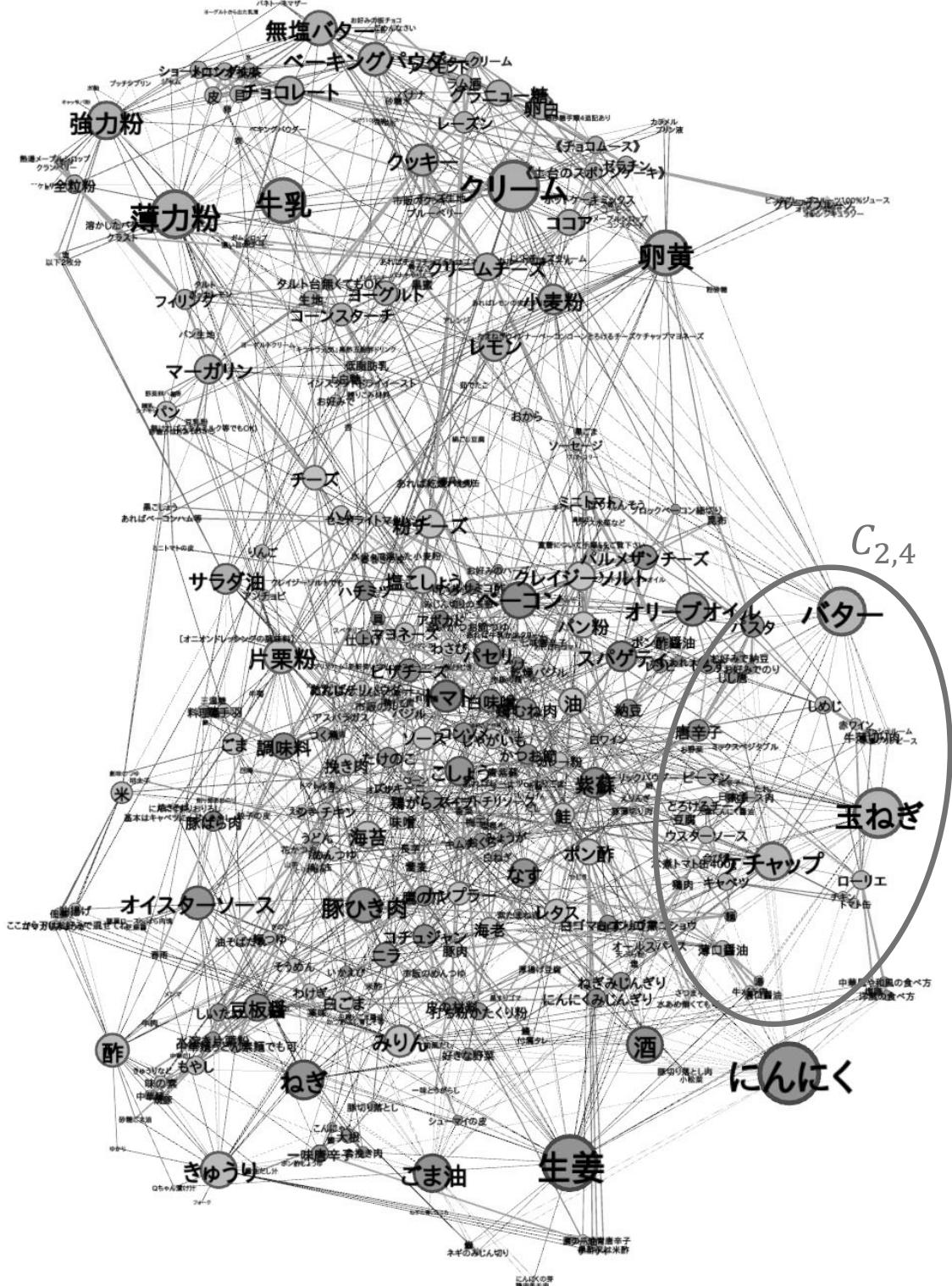


図 1 協調グループ 2 の ingredient complement network 可視化結果

その共有イベントとみなした。本研究における分析では、Cookpad の 2007 年 7 月 1 日から 12 月 31 日のデータを用いた。なお、最初の一か月間に 5 回以上共有されたアイテムを対象とし、データ上の初日以降に発生したアイテムの共有イベントを 3000 件調べていくことによりデータセットを構築した。構築したデータセットにおける

アイテム数  $M$  (レシピ総数) は 465 であった。各データセットに対して CHP モデルを適用して協調グループを抽出したところ、 $K=15$  の協調グループが抽出された。本研究では、抽出され協調グループのうち、対応するレシピが多い 2 つの主要協調グループ  $A_2$ ,  $A_4$  に着目する。

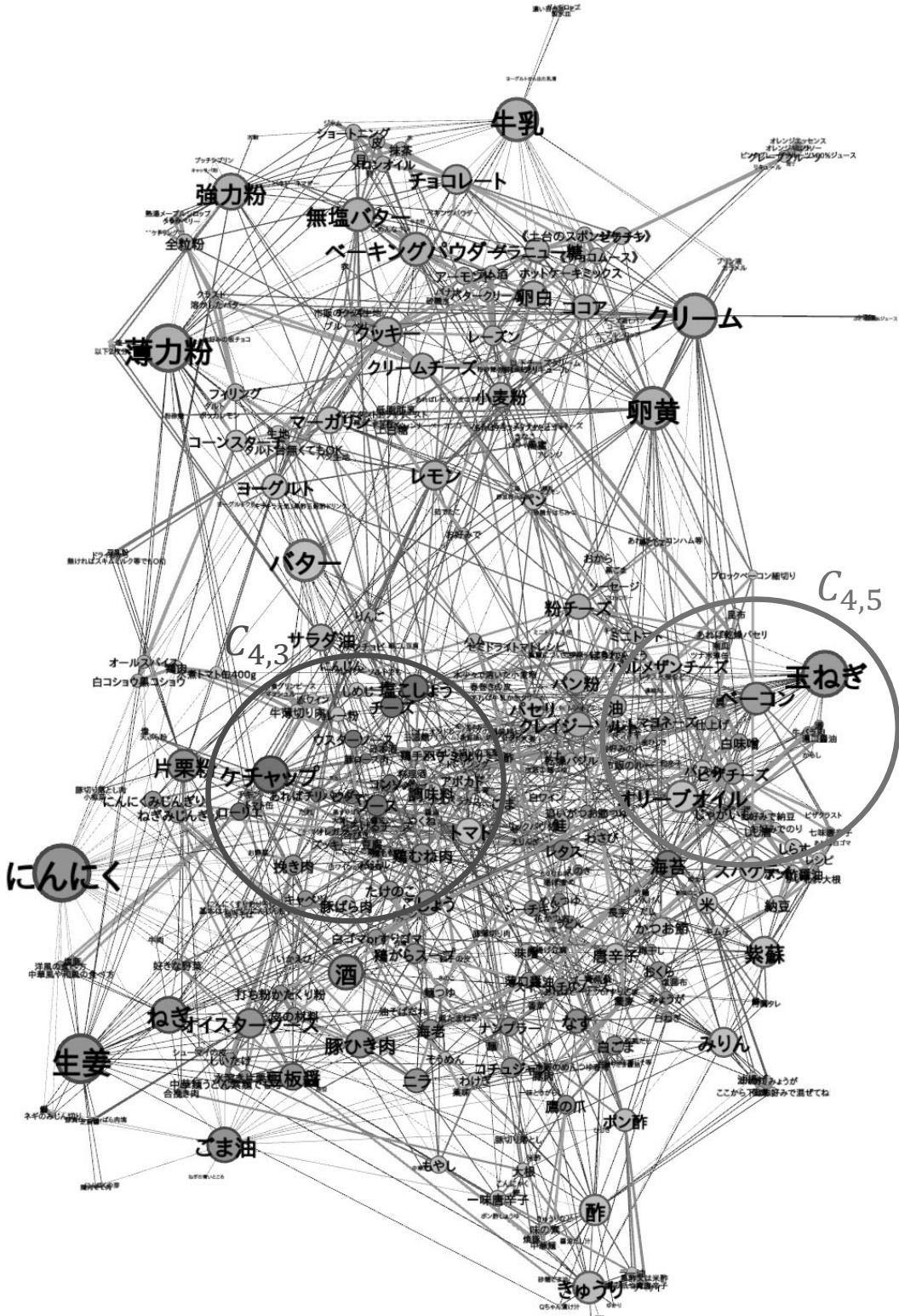


図 2 協調グループ 4 の ingredient complement network 可視化結果

#### 4.2 食材分析の結果

図 1 と図 2 に、主要協調グループ  $A_2$ ,  $A_4$  に対して構築した ingredient complement network の可視化結果を示す。ノードは食材を表し、円で描画されている。また、円の中央に食材名を記載している。食材ノード

の円の大きさは食材ノードの次数を表し、円が大きいほど次数が高いことを示している。円の色は Rosvall と Bergstrom [13] のネットワーククラスタリング法を適用して抽出されたクラスタを表しており、同じ色の円は同じクラスタに所属していることを表している。主要協調

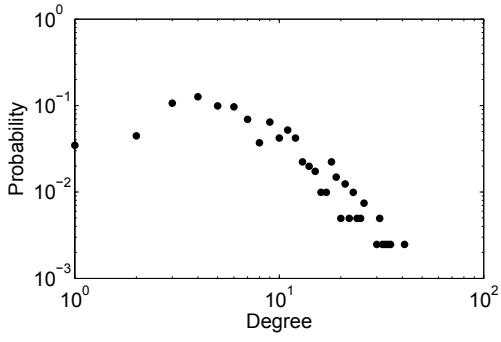
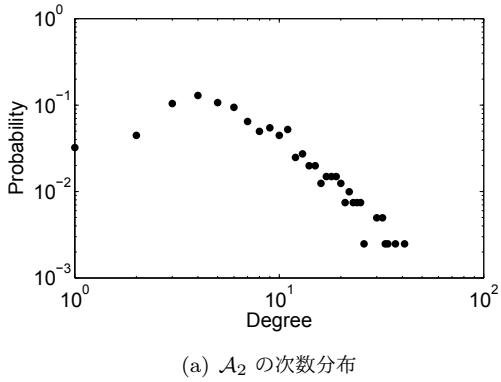


図3 主要協調グループの食材ノード次数分布

表1 次数の大きい食材

	主要協調グループ $\mathcal{A}_2$			主要協調グループ $\mathcal{A}_4$		
	食材名	次数	$C_{2,i}$	食材名	次数	$C_{4,j}$
1	にんにく	41	2	にんにく	41	2
2	生姜	37	2	生姜	35	2
3	クリーム	34	1	薄力粉	34	1
4	玉ねぎ	33	4	玉ねぎ	33	5
5	薄力粉	32	1	クリーム	32	1
6	バター	32	4	牛乳	31	1
7	牛乳	30	1	卵黄	31	1
8	卵黄	30	1	バター	30	8
9	ねぎ	26	2	強力粉	26	1
10	強力粉	25	1	ねぎ	26	2
11	ごま油	25	2	ケチャップ	26	3
12	ケチャップ	25	4	ごま油	25	2
13	酒	24	2	酒	25	2
14	ベーコン	24	6	ベーキングパウダー	24	1
15	きゅうり	24	9	無塩バター	24	1

グループを食材の観点から分析するうえで、これら2つの食材ネットワークの特徴の違いを調べるために、まず、各食材ノードの次数に着目して比較分析を行う。

図3に、主要協調グループ  $\mathcal{A}_2$  と主要協調グループ  $\mathcal{A}_4$  の次数分布を示す。いずれの次数分布も、power law tail を示すことがわかる。表1は、主要協調グループ  $\mathcal{A}_2$

表2 協調グループ  $\mathcal{A}_2$  の  $C_{2,4}$  とその最も近い協調グループ  $\mathcal{A}_4$  のクラスタ ( $C_{4,3}$ )

	$C_{2,4}$ ( $C_{4,3}$ と共に)	$C_{4,3}$ ( $C_{2,4}$ と共に)
1	ケチャップ	ケチャップ
2	ローリエ	ローリエ
3	ピーマン	ウスターソース
4	ウスターソース	とろけるチーズ
5	とろけるチーズ	ピーマン
6	豆腐	豆腐
7	合びき	合びき
8	日本酒	日本酒
9	豚ロース肉	豚ロース肉
10	チキン	トマト缶
	$C_{2,4}$ のみに含まれた食材	$C_{4,3}$ のみに含まれた食材
1	玉ねぎ	塩こしょう
2	バター	チーズ
3	キャベツ	ソース
4	しめじ	コンソメ
5	薄口醤油	あればチリパウダー
6	牛薄切り肉	ズッキーニ
7	麺	オレガノ
8	赤ワイン	杏
9	牛バラ肉	お野菜
10	湯	

と  $\mathcal{A}_4$  それぞれの食材ネットワークにおいて次数の高い Top 15 食材を示したものである。いずれの主要協調グループにおいても順位が同じとなった同一食材は、上位から「にんにく」、2位が「生姜」、4位が「玉ねぎ」となっている。また、表1では、各食材が所属するクラスタ番号  $i, j$  を記載しており、 $C_{2,i}$  は、主要協調グループ  $\mathcal{A}_2$  の  $i$  番目のクラスタ  $C_{4,j}$  は、主要協調グループ  $\mathcal{A}_4$  の  $j$  番目のクラスタを表している。なお、クラスタ番号  $i, j$  は、クラスタに含まれる食材数が多い順に1から始まる番号を割り当てている。この観点において主要協調グループ  $\mathcal{A}_2$  と  $\mathcal{A}_4$  の上位 5 位までを比較すると、主要協調グループ  $\mathcal{A}_2$  で最も食材数の多いクラスタ  $C_{2,1}$  に次数ランキング 3 位と 5 位の「クリーム」と「薄力粉」が所属しており、クラスタ  $C_{2,2}$  に次数ランキング 1 位と 2 位の「にんにく」と「生姜」が所属している。この所属構成は、主要協調グループ  $\mathcal{A}_4$  と全く同じになることがわかる。ただし、主要協調グループ  $\mathcal{A}_2$  ではクラスタ  $C_{2,4}$  に「玉ねぎ」が所属しているが、主要協調グループ  $\mathcal{A}_4$  では食材数が 5 位のクラスタ  $C_{2,5}$  に所属しているため、順位が異なっていることがわかる。

#### 4.3 食材ネットワークのクラスタ分析の結果

次に、主要協調グループ  $\mathcal{A}_2$  と  $\mathcal{A}_4$  のクラスタ間類似性について分析する。そこで、主要協調グループ  $\mathcal{A}_2$  の各クラスタ  $C_{2,i}$  が、jaccard 係数に基づいて主要協調グループ  $\mathcal{A}_4$  のどのクラスタ  $C_{4,j}$  と最も似ているかを求めた。ここでは、頁数の関係で主要協調グループ  $\mathcal{A}_2$  と  $\mathcal{A}_4$  でクラスタの食材数ランキングにおいて所属するク

表 3 協調グループ  $A_4$  のクラスタ ( $C_{4,5}$ )

1	玉ねぎ	8	具
2	ベーコン	9	市販のルー
3	ピザチーズ	10	あれば乾燥パセリ
4	マヨネーズ	11	南瓜
5	じゃがいも	12	七味唐辛子
6	白味噌	13	湯
7	仕上げ	14	お好みのハーブ

ラスタの順位が異なった「玉ねぎ」を含むクラスタに着目した分析結果のみを報告する。主要協調グループ  $A_2$  で「玉ねぎ」が所属したクラスタ  $C_{2,4}$ （図 1 の楕円を参照）とその最も近い主要協調グループ  $A_4$  のクラスタ  $C_{4,3}$ （図 2 の  $C_{4,3}$  の楕円を参照）に含まれる食材群に関する次数ランキングを表 2 に示す。クラスタ  $C_{2,4}$  と  $C_{4,3}$  の jaccard 係数は 0.378 であった。なお、表 2 の左側は  $C_{2,4}$  に、右側は  $C_{4,3}$  に所属する食材群であり、上部は、クラスタ  $C_{2,4}$  と  $C_{4,3}$  で共通する次数 Top 10 の食材群、下部は、それぞれのクラスタのみに所属する次数 Top 10 の食材群をそれぞれ示している。 $C_{2,4}$  で最も次数が高かった「玉ねぎ」は、表 2 の  $C_{4,3}$  に無いことがわかる。実際に、「玉ねぎ」は  $C_{4,3}$  には所属していなかった。主要協調グループ  $A_4$ において「玉ねぎ」が所属するクラスタは  $C_{4,5}$ （図 2 の  $C_{4,5}$  の楕円を参照）であった。なお、 $C_{2,4}$  と  $C_{4,5}$  の jaccard 係数は 0.109 であった。ここで、クラスタ  $C_{4,5}$  の食材に関する次数ランキングを表 3 に示す。 $C_{2,4}$  と  $C_{4,5}$  の違いを表 2 と表 3 により比較すると、 $C_{2,4}$  の共通食材に「豆腐」や「日本酒」があることや  $C_{2,4}$  のみに含まれる食材に「薄口醤油」など和の食材が目立つのに対し、 $C_{4,5}$  の次数ランキング上位では、「ベーコン」や「ピザチーズ」、「マヨネーズ」など、イタリア風料理との関連が予想される食材が目立つことから、「玉ねぎ」は、主要協調グループ  $A_2$  と  $A_4$  でそれぞれ異なる料理ジャンルで利用される傾向があることが示唆される。以上から、主要協調グループどうしは大局的に同様の傾向を示すものの、クラスタの比較を行うと、特色の違いが現れる場合があることがわかった。

## 5まとめ

本研究では、アテンションダイナミクスに基づいた料理レシピ群の協調構造の特徴を明らかにすることを目指し、CHP モデルにより抽出された協調グループのうち、多くのレシピと対応している主要協調グループに所属するレシピについて、食材ネットワークを構築し、食材ノードの次数の観点から代表的な利用食材や食材ネットワークのクラスタ構造について分析を行った。今後は、旬との関係や、協調グループ間の影響構造について詳細に分析する予定である。

## 謝辞

本研究では、クックパッド株式会社と国立情報学研究所が提供するクックパッドデータを利用した。

## 参考文献

- [1] Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P., and Barabási, A.-L., "Flavor network and the principles of food pairing," *Scientific Reports*, vol. 1, pp. 196:1–196:7, 2011.
- [2] Teng, C.-Y., Lin, Y.-R., and Adamic, L. A., "Recipe recommendation using ingredient networks," in *Proceedings of WebSci'12*, 2012, pp. 298–307.
- [3] Jain, A., N K, R. and Bagler, G., "Analysis of food pairing in regional cuisines of India," *PLoS ONE*, vol. 10, vo. 10, pp. 1–17, 2015.
- [4] Jiang, Y., Skufca, J. D. and Sun, J., "BiFold visualization of bipartite datasets," *EPJ Data Science*, vol. 6, pp. 2:1–2:19, 2017.
- [5] 菊地悠樹, 熊野雅仁, 木村昌弘, "料理レシピ共有サイトにおける食材のアクティビティ共起パターンの抽出," 情報処理学会論文誌「数理モデル化と応用」, vol. 11, no. 2, pp. 30–40, 2018.
- [6] Szabo, G. and Huberman, B., "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010.
- [7] Shen, H., Wang, D., Song, C., and Barabási, A.-L., "Modeling and predicting popularity dynamics via reinforced Poisson processes", in *Proceedings of AAAI'14*, 2014, pp. 291–297.
- [8] Zhao, Q., Erdogan, M., He, H. Rajaraman, A., and Leskovec, J., "SEISMIC: A self-exciting point process model for predicting tweet popularity," in *Proceedings of KDD'15*, 2015, pp. 1513–1522.
- [9] 松谷貫司, 熊野雅仁, 木村昌弘, 斎藤和己, 大原剛三, 元田浩, "オンラインアイテム群における共有イベント系列に基づいた協調構造の抽出," 人工知能学会論文誌, vol. 33, no. 3, p.A-HA1-1-12, 2018.
- [10] Hawkes, A., "Spectra of some self-exciting and mutually exciting point process," *Biometrika*, vol. 58, vo. 1, pp. 83–90, 1971.
- [11] Neal, R. M., "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 249–265, 2000.
- [12] Zhou, K., Zha, H., and Song, L., "Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes," in *Proceedings of AISTATS'13*, 2013, pp. 641–649.
- [13] Rosvall, M. and Bergstrom, C., "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 4, pp. 1118–1123, 2008.
- [14] Jacomy, M., Venturini, T., Heymann, S., and Bastian, M., "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software," *PROS ONE*, vol. 9, issue. 6, e98679, pp. 1–12, 2014.