

RESEARCH

Open Access



# Transition analysis of boundary-based active configurations in temporal simplicial complexes for ingredient co-occurrences in recipe streams

Koudai Fujisawa<sup>1</sup>, Masahito Kumano<sup>1</sup> and Masahiro Kimura<sup>1\*</sup>

\*Correspondence:  
kimura@rins.ryukoku.ac.jp

<sup>1</sup> Faculty of Advanced Science and Technology, Ryukoku University, Otsu, Japan

## Abstract

Aiming at knowledge discovery for temporal sequences of cooking recipes published in social media platforms from the viewpoint of network science, we consider an analysis of temporal higher-order networks of ingredients derived from such recipe streams by focusing on the framework of simplicial complex. Previous work found interesting properties of temporal simplicial complexes for the human proximity interactions in five different social settings by examining the configuration transitions before and after triplet interaction events corresponding to 2-simplices. In this paper, as an effective extension of the previous work to the case of higher dimensional  $n$ -simplices corresponding to newly published recipes, we propose a novel method of configuration transition analysis by incorporating the following two features. First, to focus on changes in the topological structure of temporal simplicial complex, we incorporate analyzing the transitions of boundary-based configurations. Next, to focus on the temporal heterogeneity in usage activities of ingredients, we incorporate analyzing the transitions of active configurations by introducing the activity degree of configuration. Using real data of a Japanese recipe sharing site, we empirically evaluate the effectiveness of the proposed method, and reveal some characteristics of the temporal evolution of Japanese homemade recipes published in social media from the perspective of ingredient co-occurrences.

**Keywords:** Configuration transition analysis, Higher-order network, Ingredient co-occurrence network, Simplicial complex, Temporal network analysis

## Introduction

The increasing popularity of social media platforms for sharing cooking recipes is enabling us to investigate creative homemade recipes of ordinary people. Recently, the interest in food science and computing has been growing (Min et al. 2019), and the data analysis methods from network science have also been used to explore the co-occurrence properties of ingredients in recipes (Ahn et al. 2011; Teng et al. 2012; Jain et al. 2015). Here, networks provide a fundamental framework for modeling complex systems, and network-based analysis has been successfully applied to various areas including social media analytics and behavioral and social sciences (Barabási 2016). In traditional network-based models, graphs have been widely used, where vertices represent the elementary units of the underlying system, and edges encode their pairwise interactions and relationships. However, in many real-world systems, it can be important to analyze interactions among more than two units (Benson et al. 2019). For instance, people mostly communicate in groups, more than two species can interact in an ecosystem, social topics can often be explained by multiple keywords, and a cooking recipe usually consists of more than two ingredients. For modeling such higher-order interactions, attention has recently been devoted to using *hypergraphs* and *simplicial complexes* since traditional graphs cannot give faithful representations (Preti et al. 2021). Hypergraphs are one of the most natural generalizations of graphs, where the concept of an edge is generalized to a *hyperedge* described as a subset of vertices. Simplicial complexes are a special case of hypergraphs, satisfying the *downward closure* property (Preti et al. 2021), where the generalization of an edge is called a *simplex*. Theory of simplicial complexes has a close relation with algebraic topology (Croom 2007), and modeling frameworks based on simplicial complexes have been successful in developing new insights into several research problems including brain organization (Saggar et al. 2018), protein interaction (Estrada and Ross 2018) and social influence spreading (Iacopini et al. 2019). In this paper, by focusing on the framework of a temporal simplicial complex, we consider the dynamical properties of higher-order relationships among ingredients appearing within a recipe stream, consisting of recipes with time-stamps, that is generated through diverse user interactions on a social media platform dedicated to sharing homemade recipes (see the “[Datasets and experimental settings](#)” section in the “[Experiments](#)” section for more details on such recipe stream data).

For an analysis of growing simplicial complexes, Benson et al. (2019) explored a higher-order link prediction problem for 19 datasets from various domains in terms of *simplicial closure*, and revealed fundamental differences between higher-order networks (i.e., simplicial complexes) and traditional dyadic networks (i.e., traditional graphs). Unlike link prediction such as simplicial closure, Cencetti et al. (2021) found interesting properties of the temporal simplicial complexes derived from human proximity interactions for five different social settings in terms of configuration transitions. More specifically, for each event corresponding to a 2-simplex (i.e., an interaction among exactly three people), they analyzed the configurations among the three people involved one step before the interaction and one step after it, and revealed the characteristics for transitions of the configurations. However, in the case of  $n$ -simplices with  $n \geq 3$ , such configuration transition analysis was not performed. In fact, this analysis method can intrinsically suffer from combinatorial explosion, i.e., the number of possible configurations can become

very large as the dimension  $n$  of simplices to be considered gets increasingly larger. Thus, it is desirable to develop an effective and efficient approach for configuration transition analysis.

In this paper, we effectively extend the previous work to higher dimensional  $n$ -simplices, and consider the problem of analyzing configuration transitions for a temporal simplicial complex of ingredient co-occurrences derived from a recipe stream. Here, first assuming that nice homemade foods can be made through fascinating combinations of ingredients, we focus on each combination of the  $n$  ingredients for a recipe consisting of  $(n + 1)$  ingredients as a first-step. Next, assuming that ingredients used in recipes can strongly depend on season and trend, we also take temporal usage patterns of ingredients into account. We thus present a novel method of configuration transition analysis from the following perspectives: First, aiming at effectively capturing changes in the basic qualitative features of ingredient combinations in recipes, we focus on changes in the topological structures of simplicial complexes before and after occurrences of  $n$ -simplices corresponding to newly published recipes, and propose analyzing the transitions of boundary-based configurations according to *homology theory* (Croom 2007) from algebraic topology, instead of examining all possible configurations. Next, for the  $n$ -simplices corresponding to new recipes, we propose analyzing the transitions of active configurations around them by introducing the activity degree of configuration to focus on the temporal heterogeneity in usage activities of ingredients in the recipe stream. Using real data from a social media site for sharing Japanese recipes, we empirically evaluate the proposed method for configuration transition analysis, and extract characteristics of the temporal evolution of Japanese homemade recipes in terms of ingredient combinations.

This paper extends our conference paper (Fujisawa et al. 2023) in which we presented the basic idea for analyzing the transitions of boundary-based active configurations between the preceding and forthcoming periods for occurrences of  $n$ -simplices, and showed the analysis results for a part of the dataset of Japanese recipe sharing site “Cookpad”<sup>1</sup> only in the cases of 3-simplices (i.e.,  $n = 3$ ) and 4-simplices (i.e.,  $n = 4$ ) by setting one month as the preceding and forthcoming periods. In this paper, we first provide the explanation of our configuration transition analysis method in more detail. By improving the analysis method, we further propose such a framework that is able to effectively analyze the whole picture of the configuration transitions around  $n$ -simplices over all dimensions  $n$ . We also evaluate the proposed analysis framework in other parts of the Cookpad dataset instead of repeating the experimental evaluation for the same dataset as in the conference paper (Fujisawa et al. 2023). To show the significance of incorporating the activity degree of configuration, we newly compare it with a baseline method derived as a straightforward extension of the previous work (Cencetti et al. 2021). Moreover, we examine the effect of the length of the preceding and forthcoming periods, which is an important parameter for analysis.

The paper is organized as follows. The “[Related work](#)” section briefly mentions related work. As preliminaries, the “[Preliminaries](#)” section briefly describes the previous work (Cencetti et al. 2021) for configuration transition analysis of human social interactions,

---

<sup>1</sup> <https://cookpad.com>/<https://cookpad.com/>

explains the temporal simplicial complexes derived from recipe streams, and introduces several notations used in the later sections. The “[Analysis framework](#)” section explains the proposed analysis framework in detail, and the “[Experiments](#)” section reports the experimental results for real data. The “[Conclusion](#)” section summarizes the main achievement and future plans.

### **Related work**

Food science and computing have been attracting attention in recent years (Min et al. 2019). As an analysis from the perspective of complex network science, pairwise relationships between ingredients or cuisines were explored. Several researchers investigated dyadic networks of ingredients for various cuisines in terms of flavor compounds (Ahn et al. 2011; Jain et al. 2015; Park et al. 2021; Makinei and Hazarika 2022). For example, Jain et al. (2015) analyzed ingredient networks in terms of spices for Indian cuisines. To make a comparative analysis of cuisines in the world, Sajadmanesh et al. (2017) examined cuisine networks in terms of ingredients and flavors. On the other hand, West et al. (2013) investigated population-wide dietary preferences from recipe queries on the Web, Jiang et al. (2017) compared food cultures in the world by jointly visualizing recipe density and ingredient categories, and Min et al. (2018) made a cross-region analysis of culinary culture by automatically extracting cuisine-course topics of recipes on the basis of ingredient combinations. Also, a variety of food-oriented applications are being explored. For example, complement and substitute networks of ingredients were analyzed toward recipe recommendation (Teng et al. 2012), and it was investigated how algorithmic solutions relate to the healthiness of recipes (Trattner and Elswailer 2017). However, as far as we know, there have been few attempts to analyze higher-order relationships among ingredients for home recipes. In this paper, we explore the characteristics for the temporal evolution of higher-order relationships among ingredients for Japanese homemade recipes published in social media.

This paper is also related to link prediction tasks. Much effort has been devoted to studying link prediction for traditional dyadic-networks (Liben-Nowell and Kleinberg 2007; Leskovec et al. 2010; Lu and Zhou 2011). However, only a few attempts have been made at higher-order link prediction for large real-world data since it can be computationally challenging. For example, Xu et al. (2013) provided HPLSF which is a method of predicting higher-order links of an arbitrary order for social networks using latent features. By restricting a set of candidate higher-order links, Zhang et al. (2018) presented Coordinated Matrix Minimization (CMM) which is a higher-order link prediction method based on adjacency space, and empirically showed that CMM outperforms several baselines including HPLSF. As mentioned before, to reduce the computational load of analyzing growing simplicial complexes, Benson et al. (2019) focused on simplicial closure phenomena and examined higher-order link prediction in terms of simplicial closure. On the other hand, there is an increasing interest in analyzing simplicial complexes from the perspective of data science. In fact, such an analysis framework has been successfully applied to various real problems (Saggar et al. 2018; Estrada and Ross 2018; Iacopini et al. 2019), and novel generalizations of methods for traditional graphs are being developed for simplicial complexes. For instance, Shaub et al. (2020) studied diffusion processes, random walks and Laplacians on simplicial complexes, and Preti

et al. (2021) developed an efficient algorithm for truss decomposition of simplicial complexes. Also, random simplicial complex models have been mathematically investigated (Bobrowski and Krioukov 2022). In this paper, as with Cencetti et al.'s work (2021), we analyze the configuration transitions in a temporal simplicial complex that are associated with new higher-order events.

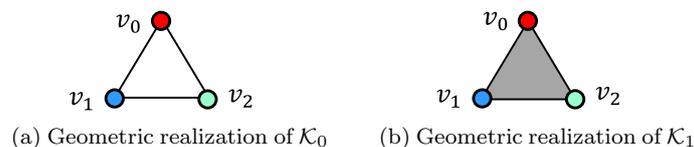
**Preliminaries**

We consider a recipe stream  $\mathcal{R}$  from a social media platform of sharing recipes, where  $\mathcal{R}$  is a set of recipes with time-stamps. For each recipe  $r \in \mathcal{R}$ , let  $t(r)$  denote the time-stamp of  $r$ . Here, we measure  $t(r)$  by using one day as unit of time. We explore the evolution of higher-order relationships among ingredients obtained from the discrete-time event history  $\mathcal{R}$  through a configuration transition analysis before and after the publication of a new recipe in  $\mathcal{R}$ .

**Previous work**

We begin with revisiting the previous work (Cencetti et al. 2021) for configuration transition analysis of temporal evolution of higher-order social interactions.

Unlike link prediction tasks such as simplicial closure (Benson et al. 2019), Cencetti et al. (2021) explored the temporal evolution of the higher-order structure of human proximity interactions in terms of configuration transition. They said that a group of  $n$  individuals is formed if there are all pairwise interactions among them during some time period. Here, higher-order interactions among people happen in a group of them. For each event corresponding to the formation of a group of  $n$  individuals, they addressed the relational structures around the  $n$  people at the preceding and forthcoming periods. In particular, they focused on group formations involving exactly three individuals (i.e., triplet interaction events), and investigated the transition rates of the relational structures around three people before and after a triplet interaction event. Here, as for possible configurations around three individuals, there are the following four classes. In the first class, there are no pairwise interactions among them, i.e., there are no groups including two of them. In the second class, there is only a single pairwise interaction among them. In the third class, there are only two pairwise interactions among them. In the fourth class, there is a group including all of them. For five different social settings, they empirically demonstrated that the configuration transitions towards and from social triplet interaction events are characterized by the above second and third classes (Cencetti et al. 2021).



**Fig. 1** Example for topological structures of simplicial complexes. Consider 2-simplex  $S = \{v_0, v_1, v_2\}$ , and its boundary-faces  $S_0 = \{v_1, v_2\}$ ,  $S_1 = \{v_0, v_2\}$ ,  $S_2 = \{v_0, v_1\}$ . Suppose that simplicial complexes at time-step 0 and time-step 1 are  $\mathcal{K}_0 = \{S_0, S_1, S_2, \{v_0\}, \{v_1\}, \{v_2\}\}$  and  $\mathcal{K}_1 = \{S, S_0, S_1, S_2, \{v_0\}, \{v_1\}, \{v_2\}\}$ , respectively. Then, the first homology groups are given by  $H_1(\mathcal{K}_0) = \mathbb{Z}$  and  $H_1(\mathcal{K}_1) = \{0\}$ . Namely,  $[\mathcal{K}_0]$  has one 1-dimensional hole, while  $[\mathcal{K}_1]$  has no 1-dimensional holes. This implies that the topological structures of  $[\mathcal{K}_0]$  and  $[\mathcal{K}_1]$  are different, and  $\mathcal{K}_0$  and  $\mathcal{K}_1$  are qualitatively distinct

**Temporal simplicial complex from recipe stream**

Given a finite set of vertices  $V$ , a  $k$ -simplex  $S$  is a subset of  $V$  with  $|S| = k + 1$  for a non-negative integer  $k < |V|$ , and a *simplicial complex*  $\mathcal{K}$  is a set of simplices which satisfies the following downward closure property (see Schaub et al. 2020; Preti et al. 2021): For every simplex  $S \in \mathcal{K}$ , all subsets of  $S$  belong to  $\mathcal{K}$ . Here, for each subset  $S'$  of simplex  $S$  with  $|S'| = k' + 1$ ,  $S'$  is called a  $k'$ -face of  $S$ , and  $S$  is called a *coface* of  $S'$ . Also, for a  $k$ -simplex  $S$ ,  $k$  is referred to as the *dimension of  $S$* . Simplicial complex  $\mathcal{K}$  corresponds to a polytope  $[\mathcal{K}]$  called its *geometric realization*, and is studied in the field of algebraic topology (Croom 2007). For an integer  $k$  with  $0 < k < |V|$  and a  $k$ -simplex  $S = \{v_0, v_1, \dots, v_k\}$ , the boundary  $\partial S$  of  $S$  consists of its  $(k - 1)$ -faces

$$S_0 = \{v_1, \dots, v_k\}, S_1 = \{v_0, v_2, \dots, v_k\}, \dots, S_k = \{v_0, v_1, \dots, v_{k-1}\}$$

which are  $(k - 1)$ -simplices in  $\mathcal{K}$  (see Fig. 1). We refer to  $S_j$  as *boundary-face* of  $S$  for each  $j = 0, 1, \dots, k$ . By algebraically examining the simplices in  $\mathcal{K}$  and their boundaries, the  $k'$ th homology group of  $\mathcal{K}$ ,  $H_{k'}(\mathcal{K})$ , is defined for any integer  $k'$  with  $0 \leq k' \leq |V|$  as a global geometric structure of the polytope  $[\mathcal{K}]$  corresponding to  $\mathcal{K}$  (Croom 2007). It is known that  $H_0(\mathcal{K})$  algebraically expresses information of the connected components in  $[\mathcal{K}]$ , and  $H_{k'}(\mathcal{K})$  algebraically represents information of the  $k'$ -dimensional holes (i.e., the holes with a  $k'$ -dimensional boundary) in  $[\mathcal{K}]$  for each  $k' \geq 1$  (see Fig. 1). In this paper, for configurations around new simplices appearing in a temporal simplicial complex, we consider focusing on their boundary-faces from the viewpoint of the topological structures of simplicial complexes through the homology theory.

As for recipe stream  $\mathcal{R}$ , we choose a set of ingredients as  $V$ , and consider the set of recipes at each time-step  $t$ ,

$$\mathcal{R}_t = \{r \in \mathcal{R} \mid t(r) = t\}.$$

We focus on a simplicial complex  $\mathcal{K}_t$ , which is defined from  $\mathcal{R}_t$  through ingredient co-occurrences in the following way: A subset  $S = \{v_0, \dots, v_k\}$  of  $V$  with  $|S| = k + 1$  is called a  *$k$ -simplex at time-step  $t$*  if there is an  $r \in \mathcal{R}_t$  such that ingredients  $v_0, \dots, v_k$  are included in recipe  $r$ .  $\mathcal{K}_t$  is defined as the set of all simplices at time-step  $t$ . We refer to  $\{\mathcal{K}_t\}$  as the *temporal simplicial complex derived from recipe stream  $\mathcal{R}$* .

For an arbitrary recipe  $r \in \mathcal{R}$ , let  $\sigma(r)$  denote the set of all the ingredients included in recipe  $r$ . We refer to  $\sigma(r)$  as the *simplex corresponding to recipe  $r$* , or a *recipe simplex*. Note that  $\sigma(r) \in \mathcal{R}_{t(r)}$ . We investigate the configuration transitions associated with occurrences of recipe simplices in temporal simplicial complex  $\{\mathcal{K}_t\}$ .

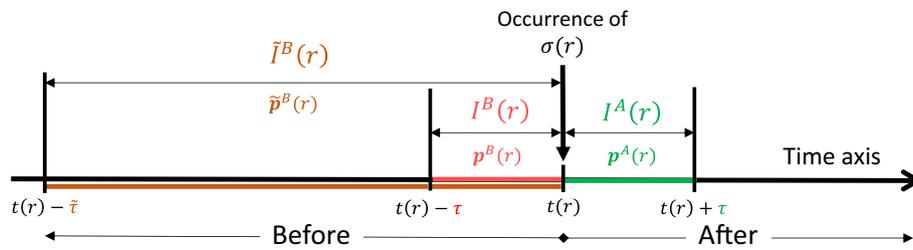
**Boundary-based configuration for recipe simplex**

For a recipe simplex

$$\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\}$$

of dimension  $n$  in temporal simplicial complex  $\{\mathcal{K}_t\}$ , we focus on a period of time just before the occurrence of  $\sigma(r)$ ,

$$I^B(r) = [t(r) - \tau, t(r) - 1]$$



**Fig. 2** Illustration of temporal change for recipe simplex  $\sigma(r)$

and a period of time just after it,

$$I^A(r) = [t(r) + 1, t(r) + \tau],$$

where a positive integer  $\tau$  is a parameter of our analysis, and indicates the length of investigation periods  $I^B(r)$  and  $I^A(r)$  (see Fig. 2). Let  $\mathcal{K}^B(r)$  and  $\mathcal{K}^A(r)$  be the simplicial complexes during periods  $I^B(r)$  and  $I^A(r)$ , respectively,

$$\mathcal{K}^B(r) = \bigcup_{t \in I^B(r)} \mathcal{K}_t, \quad \mathcal{K}^A(r) = \bigcup_{t \in I^A(r)} \mathcal{K}_t. \tag{1}$$

For temporal simplicial complex  $\{\mathcal{K}_t\}$ , we consider the configurations around  $\sigma(r)$  within investigation periods  $I^B(r)$  and  $I^A(r)$ , respectively. Note that if we straightforwardly apply the idea of the previous work (Cencetti et al. 2021) to this case, we must examine the  $k$ -faces of  $\sigma(r)$  for any integer  $k$  with  $0 \leq k < n$  and all the cofaces of  $\sigma(r)$ . To avoid this combinatorial issue, we choose not to examine all the possible configurations. Alternatively, aiming to make a connection with the topological structures of  $\mathcal{K}^B(r)$  and  $\mathcal{K}^A(r)$  through homology theory, we focus on its boundary-faces  $\sigma_0(r), \sigma_1(r), \dots, \sigma_n(r)$  for the configurations around  $\sigma(r)$  (see Fig. 3), where

$$\begin{aligned} \sigma_0(r) &= \{v_1(r), \dots, v_n(r)\}, \quad \sigma_1(r) = \{v_0(r), v_2(r), \dots, v_n(r)\}, \quad \dots, \\ \sigma_n(r) &= \{v_0(r), v_1(r), \dots, v_{n-1}(r)\}. \end{aligned}$$

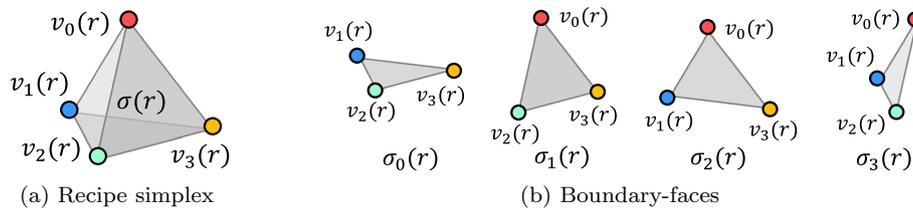
To this end, we introduce a *configuration feature vector within  $I^B(r)$*  (see Fig. 2), which is defined as

$$\mathbf{p}^B(r) = \left( p_h^B(r), p_0^B(r), \dots, p_n^B(r), p_\ell^B(r) \right),$$

and a *configuration feature vector within  $I^A(r)$*  (see Fig. 2), which is defined as

$$\mathbf{p}^A(r) = \left( p_h^A(r), p_0^A(r), \dots, p_n^A(r), p_\ell^A(r) \right).$$

Now, we define  $\mathbf{p}^B(r)$  and  $\mathbf{p}^A(r)$  in explicit detail. First,  $p_h^B(r)$  and  $p_h^A(r)$  are the probabilities that a recipe simplex including  $\sigma(r)$  occurs within  $I^B(r)$  and  $I^A(r)$ , respectively. Next, for each  $j = 0, 1, \dots, n$ ,  $p_j^B(r)$  and  $p_j^A(r)$  are the probabilities that a recipe simplex including only one  $\sigma_j(r)$  among  $\sigma_0(r), \sigma_1(r), \dots, \sigma_n(r)$  occurs within  $I^B(r)$  and  $I^A(r)$ , respectively (see Fig. 3). Finally,  $p_\ell^B(r)$  and  $p_\ell^A(r)$  are the probabilities that a recipe



**Fig. 3** Example for the boundary-faces of a recipe  $n$ -simplex  $\sigma(r)$  ( $n = 3$ )

simplex including none of  $\sigma_0(r), \sigma_1(r), \dots, \sigma_n(r)$  occurs within  $I^B(r)$  and  $I^A(r)$ , respectively. As for the probabilities described above, we empirically estimated them in our experiments.

To take into account the temporal heterogeneity in usage activities of ingredients in recipe stream  $\mathcal{R}$ , we also introduce a *concept of activity degree*. To this end, we consider a sufficiently long period of time before the occurrence of  $\sigma(r)$ ,

$$\tilde{I}^B(r) = [t(r) - \tilde{\tau}, t(r) - 1],$$

and introduce a *configuration feature vector within  $\tilde{I}^B(r)$* , which is defined as

$$\tilde{\mathbf{p}}^B(r) = (\tilde{p}_h^B(r), \tilde{p}_0^B(r), \dots, \tilde{p}_n^B(r), \tilde{p}_\ell^B(r))$$

(see Fig. 2). Here, the length  $\tilde{\tau}$  of investigation period  $\tilde{I}^B(r)$  is a positive integer with  $\tilde{\tau} \gg \tau$ , and  $\tilde{\mathbf{p}}^B(r)$  is defined within  $\tilde{I}^B(r)$  in the same way as  $\mathbf{p}^B(r)$  within  $I^B(r)$ . Note that unlike  $\mathbf{p}^B(r)$ ,  $\tilde{\mathbf{p}}^B(r)$  represents the ordinary feature for the co-occurrence of ingredients  $v_0(r), v_1(r), \dots, v_n(r)$  in recipes before time-step  $t(r)$  if  $\tilde{I}^B(r)$  is a sufficiently long period. In the experiments, we set  $\tilde{\tau}$  to 365 days (i.e., one year). We are also interested in how configuration feature vectors  $\mathbf{p}^B(r)$  within  $I^B(r)$  and  $\tilde{\mathbf{p}}^B(r)$  within  $\tilde{I}^B(r)$  affect configuration feature vector  $\mathbf{p}^A(r)$  within  $I^A(r)$  (see Fig. 2).

### Analysis framework

For temporal simplicial complex of ingredient co-occurrences  $\{\mathcal{K}_t\}$  derived from recipe stream  $\mathcal{R}$ , we propose an analysis framework of configuration transition.

#### Transition analysis of active configurations

For each positive integer  $n$  with  $0 < n < |V|$ , let  $\mathcal{R}(n)$  denote the subset of  $\mathcal{R}$  such that  $\sigma(r)$  is an  $n$ -simplex for any  $r \in \mathcal{R}(n)$ . We first provide an analysis method of configuration transition associated with occurrence of  $\sigma(r)$  in  $\{\mathcal{K}_t\}$  for any  $r \in \mathcal{R}(n)$  under fixed  $n$ , and by appropriately improving it, we propose an analysis framework of configuration transition for the whole  $\mathcal{R}$ .

For any recipe  $n$ -simplex  $\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\}$  ( $r \in \mathcal{R}(n)$ ), we explore boundary-based configurations around  $\sigma(r)$  within investigation periods  $I^B(r)$  and  $I^A(r)$  by measuring the activity degree of configuration on the basis of the configuration feature vectors  $\mathbf{p}^B(r)$ ,  $\mathbf{p}^A(r)$  and  $\tilde{\mathbf{p}}^B(r)$ . First, for  $I^B(r)$  which is the time period just before the occurrence of  $\sigma(r)$ , we consider the decomposition of  $\mathcal{R}(n)$ ,

$$\mathcal{R}(n) = \mathcal{R}_h^B(n) \cup \bigcup_{\lambda=1}^{n+1} \mathcal{R}_\lambda^B(n) \cup \mathcal{R}_\ell^B(n) \quad (\text{disjointunion}),$$

and define the boundary-based active configuration  $\mathcal{C}(\sigma(r); I^B(r))$  around  $\sigma(r)$  within  $I^B(r)$  as follows: If  $p_h^B(r) > \tilde{p}_h^B(r)$ , then  $r \in \mathcal{R}_h^B(n)$  and we say that the boundary-based active configuration around  $\sigma(r)$  within  $I^B(r)$  is  $\sigma(r)$  itself, i.e.,  $\mathcal{C}(\sigma(r); I^B(r)) = \{\sigma(r)\}$ . Note that  $r \in \mathcal{R}_h^B(n)$  means that simplices including  $\sigma(r)$  actively occur within  $I^B(r)$ , and thus  $\sigma(r)$  itself is active within  $I^B(r)$ . In the case of  $r \notin \mathcal{R}_h^B(n)$ , we examine the inequality  $p_j^B(r) > \tilde{p}_j^B(r)$  for each boundary-face  $\sigma_j(r)$  ( $j = 0, 1, \dots, n$ ), and set

$$\mathcal{C}(\sigma(r); I^B(r)) = \{\sigma_j(r) \mid (j = 0, 1, \dots, n) \mid p_j^B(r) > \tilde{p}_j^B(r)\}. \tag{2}$$

Let  $\lambda$  denote the number of the boundary-faces belonging to  $\mathcal{C}(\sigma(r); I^B(r))$ , i.e.,  $\lambda = |\mathcal{C}(\sigma(r); I^B(r))|$ . If  $\lambda > 0$ , then  $r \in \mathcal{R}_\lambda^B(n)$  and we say that the boundary-based active configuration around  $\sigma(r)$  within  $I^B(r)$  is the set of  $\lambda$  active boundary-faces  $\mathcal{C}(\sigma(r); I^B(r))$  obtained by Eq. (2). Note that  $r \in \mathcal{R}_\lambda^B(n)$  implies that  $\lambda$  boundary-faces of  $\sigma(r)$  actively occur within  $I^B(r)$ . If  $\lambda = 0$ , then  $r \in \mathcal{R}_\ell^B(n)$  and we say that the boundary-based active configuration around  $\sigma(r)$  within  $I^B(r)$  is the empty set, i.e.,  $\mathcal{C}(\sigma(r); I^B(r)) = \emptyset$ . Note that  $r \in \mathcal{R}_\ell^B(n)$  means that none of the boundary-faces of  $\sigma(r)$  actively occur within  $I^B(r)$ .

Next, for  $I^A(r)$  which is the time period just after the occurrence of  $\sigma(r)$ , we define the decomposition of  $\mathcal{R}(n)$ ,

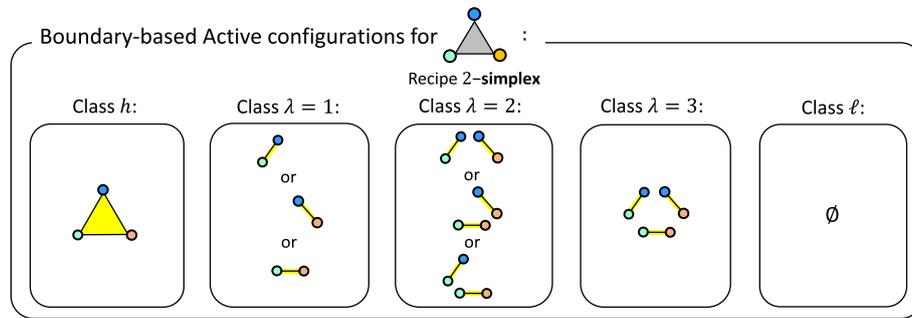
$$\mathcal{R}(n) = \mathcal{R}_h^A(n) \cup \bigcup_{\lambda=1}^{n+1} \mathcal{R}_\lambda^A(n) \cup \mathcal{R}_\ell^A(n) \quad (\text{disjointunion}),$$

and the boundary-based active configuration  $\mathcal{C}(\sigma(r); I^A(r))$  around  $\sigma(r)$  within  $I^A(r)$  in the same way as the case for  $I^B(r)$ . Here, if  $p_h^A(r) > \tilde{p}_h^A(r)$ , we define  $r \in \mathcal{R}_h^A(n)$  and  $\mathcal{C}(\sigma(r); I^A(r)) = \{\sigma(r)\}$ . In the case of  $r \notin \mathcal{R}_h^A(n)$ , we set

$$\mathcal{C}(\sigma(r); I^A(r)) = \{\sigma_j(r) \mid (j = 0, 1, \dots, n) \mid p_j^A(r) > \tilde{p}_j^A(r)\} \tag{3}$$

and  $\lambda = |\mathcal{C}(\sigma(r); I^A(r))|$ . If  $\lambda > 0$ , we define  $r \in \mathcal{R}_\lambda^A(n)$ , and say that the boundary-based active configuration around  $\sigma(r)$  within  $I^A(r)$  is the set of  $\lambda$  active boundary-faces  $\mathcal{C}(\sigma(r); I^A(r))$  obtained by Eq. (3). If  $\lambda = 0$ , we define  $r \in \mathcal{R}_\ell^A(n)$ , and  $\mathcal{C}(\sigma(r); I^A(r)) = \emptyset$ .

Therefore, for any recipe  $n$ -simplex  $\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\}$ , we divide each of boundary-based active configurations  $\mathcal{C}(\sigma(r); I^B(r))$  and  $\mathcal{C}(\sigma(r); I^A(r))$  into the following  $(n + 2)$  classes in the proposed analysis framework (see Fig. 4): First, we say that  $\mathcal{C}(\sigma(r); I^B(r))$  is in class  $h$  if  $r \in \mathcal{R}_h^B(n)$ , and  $\mathcal{C}(\sigma(r); I^A(r))$  is in class  $h$  if  $r \in \mathcal{R}_h^A(n)$ . Next, for each  $k = 1, \dots, n + 1$ , we say that  $\mathcal{C}(\sigma(r); I^B(r))$  is in class  $\lambda = k$  if  $r \in \mathcal{R}_k^B(n)$ , and  $\mathcal{C}(\sigma(r); I^A(r))$  is in class  $\lambda = k$  if  $r \in \mathcal{R}_k^A(n)$ . Finally, we say that  $\mathcal{C}(\sigma(r); I^B(r))$  is in class  $\ell$  if  $r \in \mathcal{R}_\ell^B(n)$ , and  $\mathcal{C}(\sigma(r); I^A(r))$  is in class  $\ell$  if  $r \in \mathcal{R}_\ell^A(n)$ . Note first that  $\sigma(r)$  itself is active within  $I^B(r)$  and  $I^A(r)$  if  $\mathcal{C}(\sigma(r); I^B(r))$  and  $\mathcal{C}(\sigma(r); I^A(r))$  are in class  $h$ , respectively. For each  $k = 1, \dots, n + 1$ ,  $\sigma(r)$  has  $k$  active boundary-faces within  $I^B(r)$  and  $I^A(r)$  if  $\mathcal{C}(\sigma(r); I^B(r))$  and  $\mathcal{C}(\sigma(r); I^A(r))$  are in class  $\lambda = k$ , respectively. Moreover,  $\sigma(r)$  has no active boundary-faces within  $I^B(r)$  and  $I^A(r)$  if  $\mathcal{C}(\sigma(r); I^B(r))$  and  $\mathcal{C}(\sigma(r); I^A(r))$  are in class  $\ell$ , respectively.



**Fig. 4** Five classes of the boundary-based active configurations for a recipe 2-simplex

For a fixed integer  $n$  with  $0 < n < |V|$ , we make a detailed analysis for the transition of boundary-based active configurations associated with occurrences of recipe  $n$ -simplices by using the *transition matrix*  $T(n)$ ,

$$T_{x,y}(n) = \frac{|\mathcal{R}_x^B(n) \cap \mathcal{R}_y^A(n)|}{|\mathcal{R}(n)|} \quad (x, y = \ell, 1, \dots, n + 1, h). \tag{4}$$

To effectively analyze the whole picture of the configuration transitions around recipe  $n$ -simplices across all dimensions  $n$ , we propose an analysis framework of using the *aggregated transition vector*

$$AT(n) = (AT_b(n), T_{h,h}(n), T_{h,\ell}(n), T_{\ell,h}(n), T_{\ell,\ell}(n)), \tag{5}$$

where

$$AT_b(n) = \sum_{\lambda, \lambda'=1}^{n+1} T_{\lambda, \lambda'}(n) + \sum_{\lambda=1}^{n+1} \{T_{\lambda,h}(n) + T_{h,\lambda}(n)T_{\lambda,\ell}(n) + T_{\ell,\lambda}(n)\}. \tag{6}$$

Note that  $AT_b(n)$  indicates the occurrence probability of recipe  $n$ -simplices around which the configuration transitions are comprehensively characterized by their active boundary-faces. We examine each component of  $AT(n)$  as a function  $n$ .

**Influence analysis**

We arbitrarily fix the dimension  $n$  of recipe simplices to be considered, and explore the relationship of configuration feature vector  $\mathbf{p}^A(r)$  within  $I^A(r)$  with configuration feature vectors  $\mathbf{p}^B(r)$  within  $I^B(r)$  and  $\tilde{\mathbf{p}}^B(r)$  within  $\tilde{I}^B(r)$  for each recipe  $n$ -simplex  $\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\}$ .

In view of situations before and after occurrence of  $\sigma(r)$ , we naturally speculate that we might have  $\mathbf{p}^A(r) \sim \mathbf{p}^B(r)$ ,  $\mathbf{p}^A(r) \sim \tilde{\mathbf{p}}^B(r)$ , or  $\mathbf{p}^A(r) \propto \mathbf{p}^B(r)/\tilde{\mathbf{p}}^B(r)$  (i.e.,  $p_h^A(r) = \nu p_h^B(r)/\tilde{p}_h^B(r)$ ,  $p_j^A(r) = \nu p_j^B(r)/\tilde{p}_j^B(r)$  ( $j = 0, 1, \dots, n$ ), and  $p_\ell^A(r) = \nu p_\ell^B(r)/\tilde{p}_\ell^B(r)$  for some  $\nu > 0$ ). In this paper, we examine the influence of  $\mathbf{p}^B(r)$  and  $\tilde{\mathbf{p}}^B(r)$  on  $\mathbf{p}^A(r)$  by modeling

$$\begin{aligned}
 p_h^A(r) &= c \left\{ p_h^A(r) \right\}^{w_h} \left\{ \tilde{p}_h^B(r) \right\}^{\tilde{w}_h}, \\
 p_j^A(r) &= c \left\{ p_j^B(r) \right\}^{w_b} \left\{ \tilde{p}_j^B(r) \right\}^{\tilde{w}_b} \quad (j = 0, 1, \dots, n), \\
 p_\ell^A(r) &= c \left\{ p_\ell^B(r) \right\}^{w_\ell} \left\{ \tilde{p}_\ell^B(r) \right\}^{\tilde{w}_\ell},
 \end{aligned}
 \tag{7}$$

where  $c > 0$ , and

$$\mathbf{w} = (w_h, \tilde{w}_h, w_b, \tilde{w}_b, w_\ell, \tilde{w}_\ell)$$

is the model parameter called the *weight vector*

By assuming a multinomial model within  $I^A(r)$  based on the probability vector  $\mathbf{p}^A(r)$  defined by Eq. (7), we estimate the weight vector  $\mathbf{w}$  from the observed data vector within  $I^A(r)$ ,

$$\mathbf{m}(r) = (m_h(r), m_0(r), m_1(r), \dots, m_n(r), m_\ell(r)),$$

for any  $\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\} \in \mathcal{R}(n)$  according to the MAP estimation framework. Here, let  $\mathcal{S}^A(r; n)$  be the set of such recipes in  $\mathcal{R}(n)$  that are published within  $I^A(r)$ , i.e.,

$$\mathcal{S}^A(r; n) = \left\{ r' \in \mathcal{R}(n) \mid t(r') \in I^A(r) \right\}.$$

Then,  $\mathbf{m}(r)$  is defined as follows:

$$m_h(r) = \left| \mathcal{S}_h^A(r; n) \right|, \quad m_j(r) = \left| \mathcal{S}_j^A(r; n) \right| \quad (j = 0, 1, \dots, n), \quad m_\ell(r) = \left| \mathcal{S}_\ell^A(r; n) \right|.$$

First,  $\mathcal{S}_h^A(r; n)$  is the set of recipes  $r' \in \mathcal{S}^A(r; n)$  such that recipe simplex  $\sigma(r')$  includes  $\sigma(r)$ , i.e.,

$$\mathcal{S}_h^A(r; n) = \left\{ r' \in \mathcal{S}^A(r; n) \mid \sigma(r') \supset \sigma(r) \right\}.$$

Next, for each  $j = 0, 1, \dots, n$ ,  $\mathcal{S}_j^A(r; n)$  is the set of recipes  $r' \in \mathcal{S}^A(r; n)$  such that  $r' \notin \mathcal{S}_h^A(r; n)$  and recipe simplex  $\sigma(r')$  includes boundary-face  $\sigma_j(r)$ , i.e.,

$$\mathcal{S}_j^A(r; n) = \left\{ r' \in \mathcal{S}^A(r; n) \setminus \mathcal{S}_h^A(r; n) \mid \sigma(r') \supset \sigma_j(r) \right\}.$$

Finally,  $\mathcal{S}_\ell^A(r; n)$  is the set of recipes  $r' \in \mathcal{S}^A(r; n)$  such that  $r' \notin \mathcal{S}_h^A(r; n)$  and  $r' \notin \mathcal{S}_j^A(r; n)$  for any  $j = 0, 1, \dots, n$ , i.e.,

$$\mathcal{S}_\ell^A(r; n) = \mathcal{S}^A(r; n) \setminus \mathcal{S}_h^A(r; n) \setminus \bigcup_{j=0}^n \mathcal{S}_j^A(r; n).$$

In the assumed generative model, the probability  $P(\mathbf{m}(r) \mid \mathbf{w})$  of observing  $\mathbf{m}(r) = (m_h(r), m_0(r), m_1(r), \dots, m_n(r), m_\ell(r))$  within  $I_A(r)$  is given by

$$P(\mathbf{m}(r) \mid \mathbf{w}) \propto \left\{ p_h^A(r) \right\}^{m_h(r)} \prod_{j=0}^n \left\{ p_j^A(r) \right\}^{m_j(r)} \left\{ p_\ell^A(r) \right\}^{m_\ell(r)}.
 \tag{8}$$

We assume a Gaussian prior for  $\mathbf{w}$  and estimate weight vector  $\mathbf{w}$  by maximizing the function

$$F(\mathbf{w}) = \mathcal{L}(\mathcal{R}(n); \mathbf{w}) - \frac{1}{2\mu^2} \|\mathbf{w}\|^2, \quad (9)$$

where  $\mu > 0$  is a hyper-parameter,  $\|\mathbf{w}\|$  stands for the Euclidean norm of  $\mathbf{w}$ , and  $\mathcal{L}(\mathcal{R}(n); \mathbf{w})$  is the log-likelihood of  $\mathcal{R}(n)$  for weight vector  $\mathbf{w}$ ,

$$\mathcal{L}(\mathcal{R}(n); \mathbf{w}) = \sum_{r \in \mathcal{R}(n)} \left\{ m_h(r) \log p_h^A(r) + \sum_{j=0}^n m_j(r) \log p_j^A(r) + m_\ell(r) \log p_\ell^A(r) \right\} \quad (10)$$

(see Eq. 8). Here, by Eq. (7), we have

$$\begin{aligned} p_h^A(r) &= c' \exp \left\{ w_h \log p_h^B(r) + \tilde{w}_h \log \tilde{p}_h^B(r) \right\}, \\ p_j^A(r) &= c' \exp \left\{ w_b \log p_j^B(r) + \tilde{w}_b \log \tilde{p}_j^B(r) \right\} \quad (j = 0, 1, \dots, n), \\ p_\ell^A(r) &= c' \exp \left\{ w_\ell \log p_\ell^B(r) + \tilde{w}_\ell \log \tilde{p}_\ell^B(r) \right\}, \end{aligned} \quad (11)$$

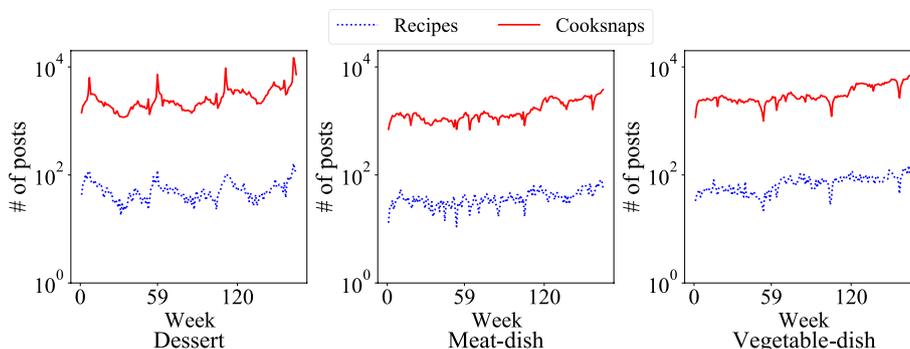
where  $c' > 0$ . Note that maximizing  $F(\mathbf{w})$  reduces to a sort of softmax optimization problem in neural network [see Eqs. (9), (10) and (11)]. In the experiments, we employed a gradient-based method (Bishop 1995).

## Experiments

We conducted an empirical evaluation of the proposed analysis framework using actual temporal simplicial complexes of ingredients obtained from Cookpad, a Japanese recipe-sharing service. Our analysis focuses on the dynamical properties of higher-order relationships involving more than two ingredients.

### Datasets and experimental settings

We utilized Cookpad data from Jan 1, 2010 to Feb 28, 2013, and constructed datasets for recipe stream  $\mathcal{R}$ , which correspond to the recipe categories of Cookpad, a social media platform dedicated to sharing Japanese homemade recipes. In particular, we focused on the three main categories of Cookpad, “Dessert”, “Meat-dish” and “Vegetable-dish”, and explored the temporal evolution of higher-order relationships among ingredients for recipes appearing within  $\mathcal{R}$ . In Cookpad, users can post their original recipes. Also, for each posted recipe, other users can send a “Thank You” message with a dish’s photo when they actually cooked and loved it, and this type of message is called a *Cooksnap*. Since the number of Cooksnaps for a recipe is seen as a measure of its popularity, it is natural for Cookpad users to aspire to create their own original recipes that receive a large number of Cooksnaps. Thus, they should try to create unique and fascinating recipes by taking into account seasonal variations and current trends. Moreover, when they come across a recipe that has received several Cooksnaps, they may be inspired to create similar ones. Consequently, it is considered that Cookpad users are likely to interact in this manner. To confirm the presence of such interactions among users for each of



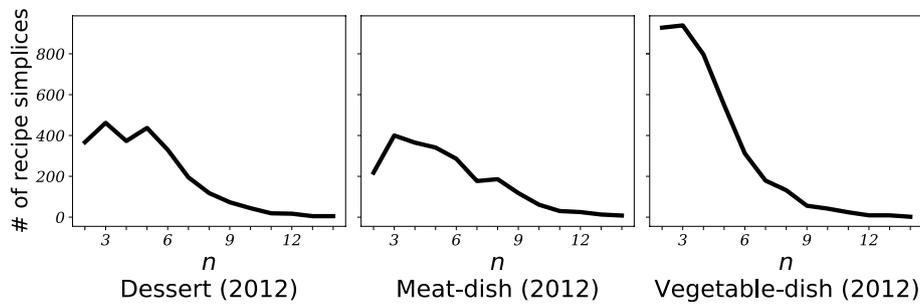
**Fig. 5** Trends in the number of recipes posted and the number of Cooksnaps received for the Dessert, Meat-dish and Vegetable-dish datasets. The number of such posts within each week is displayed for the period from Jan 1, 2010 to Feb 28, 2013

the Dessert, Meat-dish and Vegetable-dish datasets, we examined the temporal nature of recipe stream  $\mathcal{R}$ . Figure 5 shows the number of recipes posted and the number of Cooksnaps received within each week as a function of week for the three datasets. We see that the number of Cooksnaps received is much larger than the number of recipes posted for the three datasets. These results suggest that there are active interactions among Cookpad users through Cooksnaps for all three datasets. Therefore, we can assume that the three datasets for  $\mathcal{R}$  are generated through such interactions among Cookpad users, and it can be important to investigate the characteristics of recipe streams for the three datasets in terms of revealing the evolution of Japanese homemade recipes.

We separately investigated the configuration transitions for the recipes  $r$  published in 2011 and the recipes  $r$  published in 2012. In our conference paper (Fujisawa et al. 2023), we presented the analysis results for the 2011 Cookpad datasets in the cases of  $n = 3, 4$  for the dimension  $n$  of recipe simplex  $\sigma(r)$  and  $\tau = 30$  days for the length  $\tau$  of investigation periods  $I^B(r)$  and  $I^A(r)$ , while in this paper, we have conducted more detailed investigations. To avoid duplication, we only report the results for the 2012 Cookpad datasets, “Dessert (2012)”, “Meat-dish (2012)” and “Vegetable-dish (2012)”, in this paper. For each recipe stream  $\mathcal{R}$ , we adopted the set of its major ingredients<sup>2</sup> as a set of vertices  $V$ . Then, we had  $|V| = 976$ ,  $|V| = 753$ , and  $|V| = 1,023$  for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets, respectively. Fig. 6 displays the number of recipe  $n$ -simplices as a function of  $n$  for the three datasets.

On the other hand, the length  $\tau$  of investigation periods is an important parameter for our configuration transition analysis since varying the value of  $\tau$  can generally affect the emergence of active boundary-faces within  $I^B(r)$  and  $I^A(r)$  for any recipe simplex  $\sigma(r)$ . However, we only investigated the case of  $\tau = 30$  days in our conference paper (Fujisawa et al. 2023). In this paper, we examine the effect of  $\tau$  for the cases of  $\tau = 14$  days,  $\tau = 30$  days and  $\tau = 60$  days. Also, in the previous work of Cencetti et al. (2021), social interactions among three people (i.e., for the case of  $n = 2$ ) were only explored for configuration transition analysis. In this paper, we thus focus on the case of  $n \geq 3$ .

<sup>2</sup> First, general-purpose ingredients typically used in Japanese cuisines such as soy sauce, salt, sugar, water and edible oil were excluded. Next, the ingredients that appeared in at least five recipes were identified as a set of major ingredients.



**Fig. 6** Number of recipe  $n$ -simplices for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets

### Results for transition analysis of active configurations

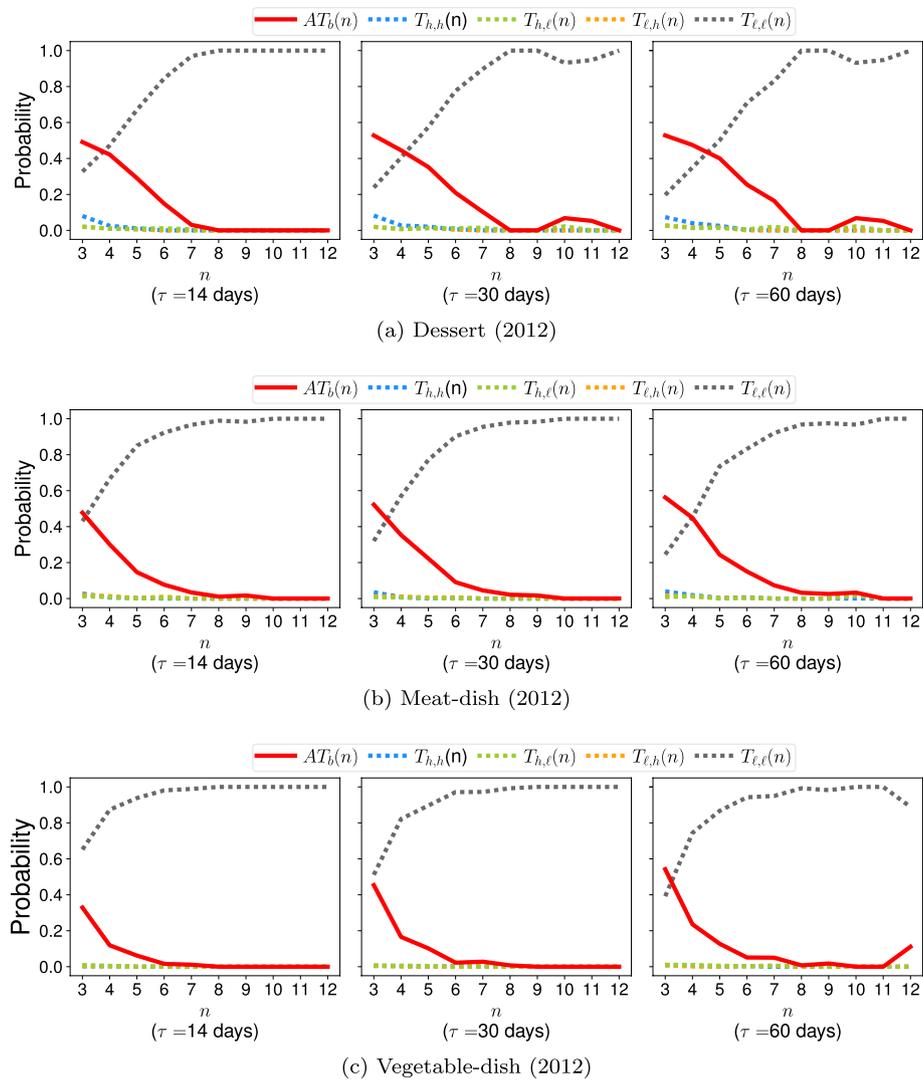
We first analyze the whole picture of the configuration transitions around recipe simplices by using the aggregated transition vector  $AT(n)$  over all dimensions  $n$  (see Eq. 6). Next, compared with a baseline method straightforwardly derived from the previous work (Cencetti et al. 2021), we describe the detailed analysis results for a particular dimension  $n$  in terms of the transition matrix  $T(n)$  (see Eq. 4).

#### Analysis results for aggregated transition vectors

For the real recipe stream  $\mathcal{R}$ , we examined the aggregated transition vectors  $\{AT(n)\}$ . Fig. 7 shows the results for the three datasets, where each component of  $AT(n)$  is displayed as a function of the dimension  $n$  of a recipe simplex. As for the length  $\tau$  of each of the investigation periods  $I^B(r)$  and  $I^A(r)$  for any recipe  $r \in \mathcal{R}(n)$  (see Fig. 2), we examined the cases of  $\tau = 14$  days,  $\tau = 30$  days and  $\tau = 60$  days.

We first see from Fig. 7 that as  $n$  increases,  $AT_b(n)$  tends to decrease and  $T_{\ell,\ell}(n)$  tends to increase. In particular,  $AT_b(n)$  and  $T_{\ell,\ell}(n)$  are dominant when  $n$  is relatively small, and  $T_{\ell,\ell}(n)$  is only dominant when  $n$  becomes larger. Note that the exact value of  $n$  that marks this shift in dominance can vary depending on the specific datasets and the value of  $\tau$ . We also observe that  $T_{\ell,h}(n)$ ,  $T_{h,\ell}(n)$  and  $T_{h,h}(n)$  are usually small. These imply that when  $n$  is relatively small, there are a significant number of recipe  $n$ -simplices around which the configuration transitions are comprehensively characterized by their active boundary-faces. However, when  $n$  is relatively large, the configuration transitions around recipe  $n$ -simplices are no longer able to be comprehensively characterized by their active boundary-faces (i.e., classes  $\lambda = 1, \dots, n + 1$ ) and themselves active (i.e., class  $h$ ). These results suggest that for such a Japanese recipe that consists of relatively small number of ingredients and is published on a social media platform, there is a high probability that some of the ingredient combinations obtained by excluding only one ingredient from the recipe can actively appear in recipes within the period before or after its occurrence. This demonstrates the effectiveness of the proposed framework for analyzing boundary-based active configurations.

Next, we consider the effect of  $\tau$ . In view of the activity degree,  $T_{\ell,\ell}(n) = 1$  and the other components of  $AT(n)$  are zero if  $\tau$  becomes equal to  $\tilde{\tau}$  (i.e., one year). On the other hand, as  $\tau$  increases, the simplicial complexes  $\mathcal{K}^B(r)$  and  $\mathcal{K}^A(r)$  for each  $r \in \mathcal{R}(n)$  (see Eq. 1) should become including more diverse simplices since there are an

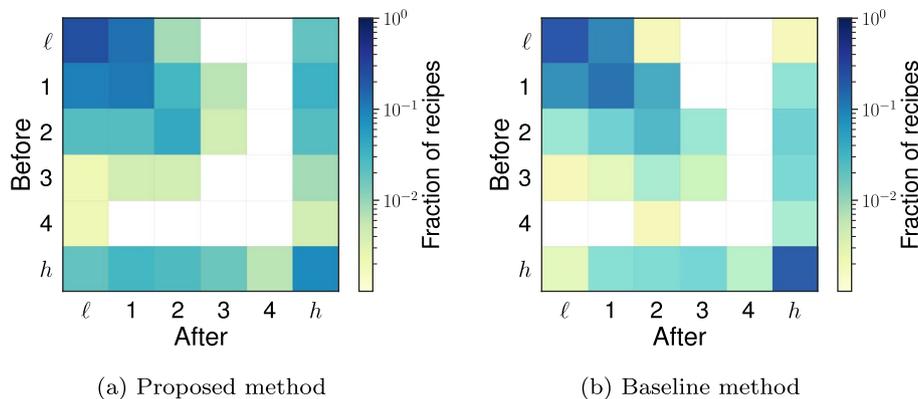


**Fig. 7** Results of aggregated transition vector  $AT(n)$  for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets. The results for  $\tau = 14$  days,  $\tau = 30$  days and  $\tau = 60$  days are displayed, where  $\tau$  is the length of each of the investigation periods  $I^B(r)$  and  $I^A(r)$  for any recipe  $r \in \mathcal{R}$

increasing number of recipes contained within  $I^B(r)$  and  $I^A(r)$ . From Fig. 7, we see that as  $\tau$  increases up to 60 days,  $AT_b(n)$  tends to increase and  $T_{l,l}(n)$  tends to decrease. Note that the magnitude of the changes observed can depend on the datasets. These results suggest that in a reasonable range of  $\tau$  such as  $14 \text{ days} \leq \tau \leq 60 \text{ days}$ , it is highly likely that the configuration transitions around recipe simplices of small dimensions can be comprehensively characterized by their active boundary-faces.

**Analysis results for transition matrices**

For the real recipe stream  $\mathcal{R}$ , we made a detailed analysis of the configuration transitions for each dimension  $n$  in terms of the transition matrix  $T(n)$ . We also evaluated the significance of incorporating the concept of activity degree through an ablation study,



**Fig. 8** Visualization results of transition matrix  $T(3)$  for the Dessert (2012) dataset ( $\tau = 30$  days)

where we consider a baseline method of only examining the presence or absence of its boundary-faces and cofaces within investigation periods  $I^B(r)$  and  $I^A(r)$  for each recipe  $n$ -simplex  $\sigma(r)$ . Note that the baseline method is obtained by ignoring the activity degree of configuration through only setting  $\tilde{p}^B(r)$  to the zero vector  $\mathbf{0}$  in the proposed analysis framework, and is regarded as a straightforward extension of the previous work (Cencetti et al. 2021). In this paper, we only report the results for the Dessert (2012) dataset in the case of  $n = 3$  and  $\tau = 30$  days. Fig. 8 shows the visualization results of transition matrix  $T(3)$ , where the results of the proposed and baseline methods are displayed in Figs. 8a and b, respectively.

As for the proposed method, we see from Fig. 8a that various entries of  $T(3)$  contributed to  $AT_b(3)$ . This implies that the configuration transitions characterized by active boundary-faces can have various types. For this dataset, those configuration transitions were mainly related to one or two active boundary-faces (i.e., classes  $\lambda = 1$  and  $\lambda = 2$ ). Thus, it is possible to extract interesting phenomena of configuration transitions in the recipe stream as illustrated later in this section. This also indicates the effectiveness of the proposed analysis framework.

From Fig. 8b, we observe that the entry  $T_{h,h}(3)$  obtained by the baseline method was large. This means that for the Dessert (2012) dataset, there were many recipe 3-simplices satisfying the condition that there exist recipe simplices including them before and after they occur. Note that this result of configuration transitions for ingredient co-occurrences is different from the properties of human interactions found by the previous work (Cencetti et al. 2021).

We also see from Figs. 8b and b that the transition matrix  $T(3)$  obtained by the proposed method was completely different from that obtained by the baseline method. For example, we focus on the entry  $T_{h,4}(3) = |\mathcal{R}_h^B(3) \cap \mathcal{R}_4^A(3)| / |\mathcal{R}(3)|$  (see Eq. 4). By the baseline method, we have a recipe  $r' \in \mathcal{R}_h^B(3) \cap \mathcal{R}_4^A(3)$  such that  $\sigma(r') = \{v_0(r') = \text{“cocoa”}, v_1(r') = \text{“butter”}, v_2(r') = \text{“chocolate”}, v_3(r') = \text{“honey”}\}$ . However, the boundary-face  $\sigma_3(r')$  of  $\sigma(r')$  is not active within  $I^A(r')$  since  $p_3^A(r') = 0.33\%$  and  $\tilde{p}_3^B(r') = 1.97\%$ . This means that it is quite usual to employ the combination of cocoa, butter and chocolate in a desert recipe. Thus, we have  $r' \notin \mathcal{R}_4^A(3)$  by the proposed method. This example clarifies the difference between the proposed and baseline methods for  $T_{h,4}(3)$ . On

the other hand, by the proposed method, we can extract a recipe  $r'' \in \mathcal{R}_h^B(3) \cap \mathcal{R}_4^A(3)$  such that  $\sigma(r'') = \{v_0(r'') = \text{“egg”}, v_1(r'') = \text{“cakeflour”}, v_2(r'') = \text{“cream cheese”}, v_3(r'') = \text{“fresh cream”}\}$ . This example reveals an interesting phenomenon as mentioned below. In fact, the configuration transition around  $\sigma(r'')$  is described as follows: Just before  $\sigma(r'')$  occurs, recipe simplices including  $\sigma(r'')$  actively occurred in simplicial complex  $\mathcal{K}^B(r'')$ . Furthermore, just after  $\sigma(r'')$  occurred, the 3-simplex  $\sigma(r'')$  decays into its four active boundary-faces  $\sigma_0(r'')$ ,  $\sigma_1(r'')$ ,  $\sigma_2(r'')$  and  $\sigma_3(r'')$  in simplicial complex  $\mathcal{K}^A(r'')$  (see Fig. 3). Namely, all the four boundary-faces actively occurred in  $\mathcal{K}^B(r'')$ , but recipe simplices including  $\sigma(r'')$  did not actively occur in  $\mathcal{K}^A(r'')$ . These results demonstrate that incorporating the concept of activity degree can be significant for exploring the ingredient co-occurrences in the recipe stream, and imply that we can find several interesting properties of temporal simplicial complex  $\{\mathcal{K}_t\}$  by examining boundary-based active configurations.

### Evaluation of influence analysis model

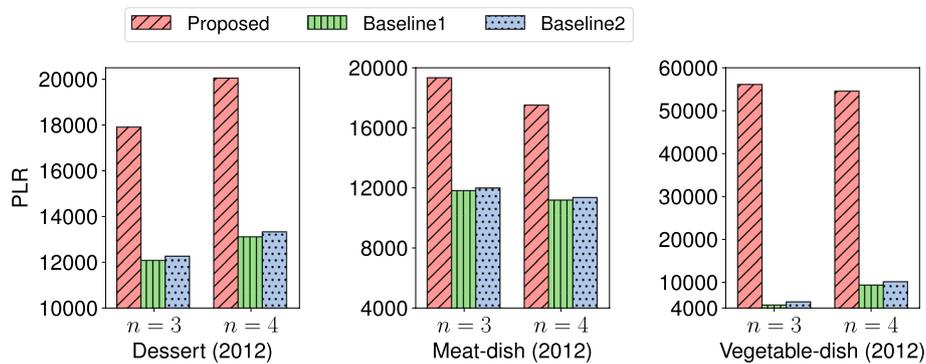
For each recipe  $n$ -simplex  $\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\}$  ( $r \in \mathcal{R}(n)$ ), we assess the influence of the configuration feature vectors  $\mathbf{p}^B(r)$  within  $I^B(r)$  and  $\tilde{\mathbf{p}}^B(r)$  within  $\tilde{I}^B(r)$  on the configuration feature vector  $\mathbf{p}^A(r)$  within  $I^A(r)$ . We first conducted an empirical evaluation of the proposed influence analysis model [see Eqs. (7) and (8)] in terms of predictive performance. We divided  $\mathcal{R}(n)$  into a training set  $\mathcal{R}^{\text{train}}(n)$  and a test set  $\mathcal{R}^{\text{test}}(n)$  along the time-axis, with a 7 : 3 ratio, and evaluated the predictive performance of the learned model on  $\mathcal{R}^{\text{train}}(n)$  using a *prediction log-likelihood ratio*  $PLR$  defined by

$$PLR = \mathcal{L}(\mathcal{R}^{\text{test}}(n); \mathbf{w}) - \mathcal{L}(\mathcal{R}^{\text{test}}(n); \mathbf{0})$$

(see Eq. 10). Here,  $PLR$  indicates the difference in log-likelihood between the learned model and a uniformly random model, and quantifies the relative performance of the learned model versus the random guessing on  $\mathcal{R}^{\text{test}}(n)$ . Note that the uniformly random model is obtained by setting  $\mathbf{w} = \mathbf{0}$ , i.e.,  $w_b = w_h = w_\ell = \tilde{w}_b = \tilde{w}_h = \tilde{w}_\ell = 0$  (see Eq. 7).

We compared the proposed generative model against four baseline models referred to as *baselines 1, 2, 3, and 4*, which are defined as follows: First, the baselines 1 and 2 are obtained by “ $w_b = w_h = w_\ell = 1, \tilde{w}_b = \tilde{w}_h = \tilde{w}_\ell = 0$ ”, and “ $\tilde{w}_b = \tilde{w}_h = \tilde{w}_\ell = 1, w_b = w_h = w_\ell = 0$ ”, respectively (see Eq. 7). Namely, the baseline 1 model supposes  $\mathbf{p}^A(r) = \mathbf{p}^B(r)$ , and the baseline 2 model supposes  $\mathbf{p}^A(r) = \tilde{\mathbf{p}}^B(r)$ . Next, the baselines 3 and 4 are obtained by “ $w_b = w_h = w_\ell = 1, \tilde{w}_b = \tilde{w}_h = \tilde{w}_\ell = -1$ ”, and “ $w_b = w_h = w_\ell = -1, \tilde{w}_b = \tilde{w}_h = \tilde{w}_\ell = 1$ ” (see Eq. 7). Namely, by considering active and inactive situations, the baseline 3 model assumes  $\mathbf{p}^A(r) \propto \mathbf{p}^B(r)/\tilde{\mathbf{p}}^B(r)$ , and the baseline 4 model assumes  $\mathbf{p}^A(r) \propto \tilde{\mathbf{p}}^B(r)/\mathbf{p}^B(r)$ , where the elementwise division of two vectors is used. In this paper, we only present the results of influence analysis for small  $n$  and  $\tau = 30$  days according to the analysis results of aggregated transition vectors (see Fig. 7).

Figure 9 indicates the evaluation results of the proposed, baseline 1 and baseline 2 models for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets in terms of the  $PLR$  metric. Here, the baseline 3 and 4 models are excluded since they were on par with the uniformly random model. First, it is evident that the proposed, baseline



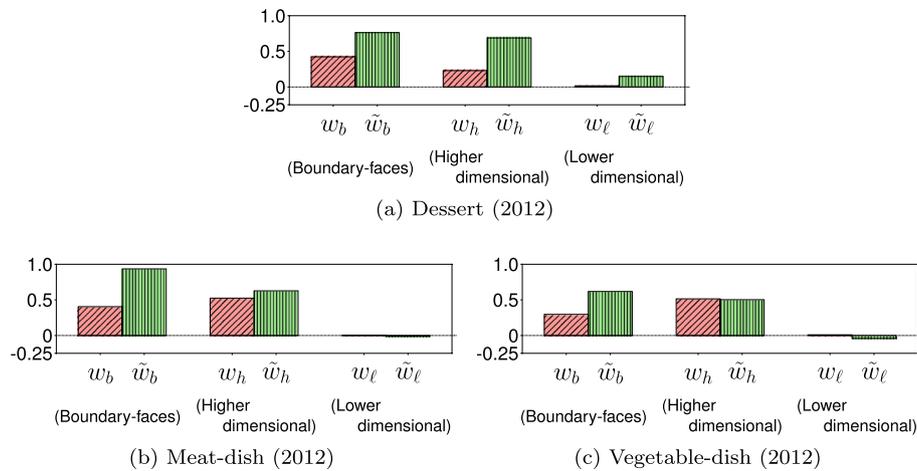
**Fig. 9** Evaluation results of the proposed influence analysis model for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets ( $n = 3, 4$  and  $\tau = 30$  days)

1 and baseline 2 models significantly outperformed the random guessing, while the value of  $PLR$  can vary depending on datasets. It should be noted that the baseline 2 model exhibited marginally better prediction performance than the baseline 1 model. However, the baseline 3 and 4 models derived from the simple ratios of  $p^B(r)$  and  $\tilde{p}^B(r)$  in view of active and inactive situations were quite ineffective. In contrast, the proposed model consistently outperformed the other models. These findings show the effectiveness of the proposed model, prompting us to utilize it for our subsequent influence analysis.

### Results for influence analysis

By applying the proposed influence analysis model [see Eqs. (7) and (8)], we examined the influence of  $p^B(r)$  and  $\tilde{p}^B(r)$  on  $p^A(r)$  for any recipe  $n$ -simplex  $\sigma(r) = \{v_0(r), v_1(r), \dots, v_n(r)\}$  ( $r \in \mathcal{R}(n)$ ) in terms of the weight vector  $w$ . In this paper, we in particular describe the analysis results for  $n = 4$  and  $\tau = 30$  days.

Figure 10 indicates the results of  $w$  for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets. Regarding the five boundary-faces  $\sigma_j(r)$  ( $j = 0, 1, \dots, 4$ ), the features  $\tilde{p}_j^B(r)$  ( $j = 0, 1, \dots, 4$ ) within long-term period  $\tilde{I}^B(r)$  had a stronger influence on  $p^A(r)$  than the features  $p_j^B(r)$  ( $j = 0, 1, \dots, 4$ ) within short-term period  $I^B(r)$  (see the results of  $w_b$  and  $\tilde{w}_b$ ). For the ingredient relationships including  $\sigma(r)$  (called *higher dimensional relationships*), the influence of the feature  $\tilde{p}_h^B(r)$  within  $\tilde{I}^B(r)$  was stronger than or comparable to that of the feature  $p_h^B(r)$  within  $I^B(r)$  (see the results of  $w_h$  and  $\tilde{w}_h$ ). Moreover, we see that which of the boundary-faces and the higher dimensional relationships had a stronger influence on  $p^A(r)$  depended on the datasets. On the other hand, as for the ingredient relationships lower than boundary-faces (called *lower dimensional relationships*), both the feature  $\tilde{p}_\ell^B(r)$  within  $\tilde{I}^B(r)$  and the feature  $p_\ell^B(r)$  within  $I^B(r)$  had a relatively weak influence on  $p^A(r)$ , compared to the other features (see the results of  $w_\ell$  and  $\tilde{w}_\ell$ ). In general, this can be attributed to the smallness of dimension  $n$ . It is worth noting that the observed properties can necessarily vary depending on the datasets to be considered. These findings suggest that the proposed influence analysis method has a potential for uncovering intriguing properties of ingredient co-occurrences in the recipe stream of social media



**Fig. 10** Results of influence analysis for the Dessert (2012), Meat-dish (2012) and Vegetable-dish (2012) datasets ( $n = 4$  and  $\tau = 30$  days)

## Conclusion

We addressed the problem of exploring configuration transitions associated with occurrences of recipe  $n$ -simplices in a temporal simplicial complex derived from a recipe stream in social media. By extending the previous work of Cencetti et al. (2021), we have proposed a novel framework of analyzing the transitions of boundary-based active configurations for the preceding and forthcoming investigation periods. First, we gave an analysis method of using the transition matrices, and by improving it, we further devised a method of employing the aggregated transition vectors so as to effectively analyze the whole picture across all dimensions  $n$ . Next, by introducing a probabilistic generative model for influence analysis, we provided a method of examining how the configuration feature vectors within the preceding periods affect the configuration feature vector within the forthcoming investigation period.

Using real data from a Japanese recipe sharing site, we empirically evaluated the effectiveness of the proposed analysis framework. We first showed that for the investigation periods of a reasonable length, the configuration transitions around recipe simplices of relatively small dimensions  $n$  are comprehensively characterized by their active boundary-faces. Next, compared with a baseline method straightforwardly derived from the previous work (Cencetti et al. 2021), we demonstrated the significance of incorporating the concept of active degree. Moreover, we revealed some interesting category-specific properties for the temporal evolution of Japanese homemade recipes published in social media from the perspective of ingredient co-occurrences.

In this paper, we focused on temporal simplicial complexes obtained from recipe streams. Clearly, the proposed framework for analyzing configuration transitions has the potential to be applied to temporal simplicial complexes for datasets from other domains. Our immediate future work is to evaluate it for other kinds of datasets. Inspired by homology theory in algebraic topology, we focused on the boundary-based active configurations before and after occurrences of simplices. Our future work

includes examining changes in the topological structures of simplicial complexes in terms of homology groups for temporal simplicial complexes in various domains.

#### Author contributions

All authors designed research, contributed new analysis methods, analyzed data, and wrote the paper.

#### Funding

This work was supported in part by JSPS KAKENHI Grant Number JP21K12152.

#### Availability of data and materials

The Cookpad dataset we used in this paper was provided by Cookpad Inc. via IDR Dataset Service of National Institute of Informatics: <https://www.nii.ac.jp/dsc/idr/cookpad/>

#### Declarations

##### Competing interests

The authors declare that they have no competing interests.

Received: 28 February 2023 Accepted: 20 July 2023

Published online: 01 August 2023

#### References

- Ahn Y-Y, Ahnert S-E, Bagrow J-P, Barabási A-L (2011) Flavor network and the principles of food pairing. *Sci Rep* 1:196–11967
- Barabási A-L (2016) *Network science*. Cambridge University Press, Cambridge
- Benson A-R, Abebe R, Schaub M-T, Jadbabaie A, Kleinberg J (2019) Simplicial closure and higher-order link prediction. *PNAS* 115(48):11221–11230
- Bishop CM (1995) *Neural networks for pattern recognition*. Oxford University Press, Oxford
- Bobrowski O, Krioukov D (2022) Random simplicial complexes: models and phenomena. In: Battiston F, Petri G (eds) *Higher-Order systems*. Springer, Cham, pp 59–96
- Cencetti G, Battiston F, Lepri B, Karsai M (2021) Temporal properties of higher-order interactions in social networks. *Sci Rep* 11:7028–1702810
- Croom F-H (2007) *Basic concepts of algebraic topology*. Springer, New York
- Estrada E, Ross G-J (2018) Centralities in simplicial complexes. applications to protein interaction networks. *J Theor Biol* 438:46–60
- Fujisawa K, Kumano M, Kimura M (2023) Analyzing configuration transitions associated with higher-order link occurrences in networks of cooking ingredients. In: *Proceedings of the 11th international conference on complex networks and their applications*, pp 623–635
- Iacopini I, Petri G, Barrat A, Latora V (2019) Simplicial models of social contagion. *Nat Commun* 9:2485
- Jain A, Nk R, Bagler G (2015) Analysis of food pairing in regional cuisines of India. *PLoS ONE* 10(10):1–17
- Jiang Y, Skufca J-D, Sun J (2017) Bifold visualization of bipartite datasets. *EPJ Data Sci* 6:2
- Leskovec J, Huttenlocher D, Kleinberg J (2010) Predicting positive and negative links in online social networks. In: *Proceedings of WWW'10*, pp 641–650
- Liben-Nowell D, Kleinberg J (2007) The link-prediction problem for social networks. *J Am Soc Inf Sci Technol* 58(7):1019–1031
- Lü L, Zhou T (2011) Link prediction in complex networks: a survey. *Phys A Stat Mech Appl* 390(6):1150–1170
- Makineji L, Hazarika M (2022) Flavour network-based analysis of food pairing: Application to the recipes of the sub-cuisines from northeast India. *Curr Res Food Sci* 5:1038–1046
- Min W, Bao B-K, Mei S, Zhu Y, Rui Y, Jiang S (2018) You are what you eat: exploring rich recipe information for cross-region food analysis. *IEEE Trans Multim* 20(4):950–964
- Min W, Jiang S, Liu L (2019) A survey on food computing. *ACM Comput Surv* 52(5):92–19236
- Park D, Kim K, Kim S, Spranger M, Kang J (2021) Flavorgraph: a large-scale food-chemical graph for generating food representations and recommending food pairings. *Sci Rep* 11(1):1–13
- Preti G, Moralest G-D-F, Bonchi F (2021) Strud: truss decomposition of simplicial ccomplexes. In: *Proceedings of WWW'21*, pp 3408–3418
- Saggar M, Sporns O, Gonzalez-Castillo J, Bandettini P-A, Carlsson G, Glover G, Reiss A-L (2018) Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nat Commun* 9:1399–1139914
- Sajadmanesh S, Jafarzadeh S, Ossia S-A, Rabiee H-R, Haddadiy H, Mejoavaz Y, Musolesi M, Cristofaro E-D, Stringhini G (2017) Kissing cuisines: exploring worldwide culinary habits on the web. In: *Proceedings of WWW'17 companion*, pp 1013–1021
- Schaub MT, Benson AR, Horn P, Lippner G, Jadbabaie A (2020) Random walks on simplicial complexes and the normalized Hodge I-Laplacian. *SIAM Rev* 62:353–391
- Teng C-Y, Lin Y-R, Adamic L-A (2012) Recipe recommendation using ingredient networks. *Proc WebSci* 12:298–307
- Trattner C, Elswiler D (2017) Implications for meal planning and recommender systems. In: *Proceedings of WWW'17*, pp 489–498

- West R, White R-W, Horvitz E (2013) From cookies to cooks: insights on dietary patterns via analysis of web usage logs. In: Proceedings of WWW'13, pp 1399–1410
- Xu Y, Rockmore D, Kleinbaum A-M (2013) Hyperlink prediction in hypernetworks using latent social features. In: Proceedings of the 16th international conference on discovery science, pp 324–339
- Zhang M, Cui Z, Jiang S, Chen Y (2018) Beyond link prediction: Predicting hyperlinks in adjacency space. In: Proceedings of AAAI'18, pp 4430–4437

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---