Recovering Reference Cooks from Directed Departures in Family Recipe Archives

Zofia Mirzelska
Centre for Information Networks
Torwica College
Gdańsk
Poland
zmirzelska@torwica.example

Danilo Frazzori
School of Archival Informatics
Instituto de Vale Manso
Braga
Portugal

Naïma El Kassari-Belot
School of Archival Informatics
Instituto de Vale Manso
Braga
Portugal

capstract:2611.00002v1 [cs.SI] 2 Nov 2026

***Abstract*—A family recipe can identify a reference cook by specifying what the contributor does differently. We test whether these departures recover independently nominated authorities in 48 private family archives devoted to potato salad. Six contributors per family supplied a card and three other relatives nominated a reference before seeing the cards. Counting distinct contributors who name a departure recovers the majority nominee in 34 families. Counting contributors who adopt a named procedure recovers the nominee in 17, and counting either relation recovers the nominee in 27. Ten of the correct departure predictions identify cooks to whom no contributor attributes an adopted procedure. The distinction matters in archives that create a name entry only for the source of an instruction actually used.**

***Index Terms*—signed networks, centrality, recipe archives, authority records**

I. Introduction

One can collect a considerable amount of information about a person’s potato salad without finding anybody who makes it. In one family archive five contributors identify the same aunt by what they have left out. The sixth gives her name as a reason to cut the potatoes while they are warm. A search limited to recipes attributed to this aunt returns no results. Three relatives interviewed separately nevertheless agree that hers is the version against which the others should be checked.

A recipe archive records several relations at once. Someone wrote down a recipe, someone supplied the current copy and there may be a third person whose opinion of the finished dish the contributor anticipates. In a searchable collection a short credit field generally serves for all three. It can represent an unqualified acknowledgment quite well. It is more difficult to enter a credit for a person known chiefly for having done the opposite. In particular, dropping the minus sign and counting every name loses the distinction between a source of instructions and a source of authority about departures from instruction.

We examine a closed collection of family recipe forums. The forums have a useful property for a study of missing accounts. They permit a contributor to identify a relative who has never posted. We can therefore keep a vertex for the person whose procedure is being discussed even if the person has no out-edges. Our target is a reference cook, the person relatives would name to settle what counts as the family version. The study makes no attempt to establish who first cooked that dish. Reference and priority can be separated without making a judgment about the quality of anybody’s salad.

The paper uses potato salad throughout. Family collections contain enough variants of this dish for a negative attribution to identify a particular cook with some precision. A writer can omit celery after a named relative, put it in against the same relative’s practice, or simply say that the salad has no celery. Only the middle form supplies evidence of a departure from that relative. These distinctions are lost at quite an early stage in a conventional graph of mentions. We preserved the clauses in which the personal names occurred and performed the projection after annotation.

II. Relations and prior work

The distinction between a tie and its interpretation has a long history in social network analysis [1], [2]. A directed edge from a recipe writer to a cook indicates an orientation of attention. It gives no sign until a relation has been specified. Signed networks are often used for relations of approval or antagonism, with higher-order constraints on the placement of negative ties [3], [4]. Departing from a parent’s recipe does not require disapproval. A cook can esteem a parent, omit that parent’s egg and name the parent in the same sentence. Imposing a balance condition would confound differences of procedure with differences between people.

We use degree rather than a recursive prestige score [5], [6]. Even a cook with no account is eligible to be a target, and several of the highest scoring targets have no outgoing links. Normalizing importance by the number of times an author is in turn cited would give an advantage to people who both host a forum and take a large part in its disputes. We record host status for an auxiliary baseline. Network boundary rules can materially change a centrality ranking even when none of the recorded ties changes [7], [8]. Removing nonmembers prior to scoring has such an effect here.

Named omissions have been described in records of crafts handed down within a household [9], [10]. A particularly informative form gives the source’s operation, then gives the author’s substitute. It creates one directed contrast although it contains two procedures. Several accounts can copy the sentence from one another. The methods for repeated senders in

Section IV limit the weight from copies by a single author but cannot in general eliminate dependence among authors. On family collections and the date assigned to a retrospective recipe, we follow Mevrini [11] and Dalverne [12]. A recipe posted yesterday may be a much older document.

III. Family forums and independent nominations

We invited moderators of invitation-only collections to open a temporary shelf devoted to one dish. Each collection had to be maintained by an extended family, allow written contributions in the study language and have at least six people willing to contribute a card. We did not inspect previous discussions. The collection retained control of its old contents, which may include claims by people who have since left the forum. Inclusion required permission to extract a small network for the temporary shelf. The search took place entirely within this material.

Sixty groups completed collection. In each, six contributors each wrote one account, with access to their own recipe cards and without seeing the other five entries until the shelf closed. We asked for the version the contributor would use if a relative wanted to make the dish. The prompt contained no request for a source or a comparison. Contributors could attach their own recipe text or type a new card. For accessibility a contributor could dictate a card to the fieldworker. The read-back ended before the worker saw any of the other entries. A post had one author in the study data regardless of who typed it.

Before writing, each group fixed a roster of eight candidate cooks. It comprised the six contributors and two additional people named by the group as cooks of the dish at some point in the family’s history. A person could be named in this way without being alive or having had a forum account. The candidates were given local identifiers. A writer could mention a ninth person, but that mention could not contribute a vote to an eligible candidate. This choice made abstention possible and gave us a stable denominator for the analysis. The list was available to annotators resolving kinship terms. Writers saw no list on the entry screen.

Three further relatives, distinct from the six contributors and the two additional cooks, each received a separate ballot. A ballot asked whose version the family would take as a reference in resolving a disagreement about how the family made potato salad. It permitted one of the eight names or a blank. Ballots were collected before relatives could read the shelf. Respondents could choose a cook whose recipe they would prefer not to eat. No one received the name of another respondent or an aggregate result. The administrator kept the nomination file apart from the text and did not release it to analysts until the classification rules had been fixed.

In 48 groups at least two voters selected the same eligible cook. The remaining twelve groups supplied either three different names or too few nonblank votes for a majority. They remain in the recruitment count. They have no uniquely defined target under the rule set before collection. Our analysis is conditional on a majority reference and describes 288 authored cards, 384 candidate positions and 144 independently solicited ballots. Counting positions prevents two relatives in different collections from merging because they have the same kinship title. We required disjoint membership from all recruited families.

IV. Coding an edge

For each pair of author and candidate we recorded whether the author explicitly adopted a procedure from the candidate and whether the author explicitly departed from one. The evidence had to identify both a procedure and the person responsible for it. “I leave out the egg” gives no contrast. “E. always added egg, which I leave out” does. “I leave out the egg as E. did” is an adoption. No sentiment dictionary can distinguish the latter two constructions by counting the word *out*. For negation with an embedded source we followed the scope tests developed by Cazerin [13] and the two-stage identity rules of Saavare [14].

A pair could carry both kinds of edge. On one card the writer accepts another cook’s choice of potato and rejects that cook’s insistence on onion. We keep a one in each adjacency matrix. Repeating the onion claim three times supplies only one contrast. Explicitly self-attributed reversals, such as having given up an ingredient on moving house, have no edge to another cook. Mentions in signatures and in a title copied automatically from the name of the shelf have no edge either. The eligibility rule was applied before resolving any ambiguity about the direction of a relation.

Two readers independently coded every card. Both knew the local family dictionary but neither had access to nominations. Readers kept the shortest span of text sufficient to justify an edge. The adjudicator saw the two spans together, which allowed a check that apparently divergent labels did not refer to different operations on the same card. Of the 2,016 eligible directed writer-candidate pairs after removing six self-pairs per group, readers differed on 73 adoption bits and 81 departure bits before adjudication. These denominators include many unmentioned relatives. A percentage agreement would be large even for readers who consistently disagreed about the hardest named mentions [15].

For candidate v let a_{uv} and d_{uv} denote the adjudicated adoption and departure indicators supplied by author u. The two degree scores and the unsigned comparison in Eq. (1) are evaluated on the same set of authors. Using the maximum for the unsigned score makes the author the unit of endorsement for all three rankings.

$$A_v = \sum_{u} a_{uv}, \qquad D_v = \sum_{u} d_{uv}, \qquad M_v = \sum_{u} \max(a_{uv}, d_{uv}). \tag{1}$$

A rule returns the sole maximizer if its largest score is positive and unique. It otherwise abstains. We count an abstention as a miss in the primary outcome. In particular, the all-zero graph cannot accidentally receive credit for placing the refer-

ence first in an alphabetical sort. Family dictionaries use identifiers in random order anyway. A secondary measure divides successes by issued predictions. We report this separately because a rule that predicts only in one large unanimous family can achieve a perfect conditional fraction.

V. Retrieval results

Counting authors who name a departure retrieves the independent reference in 34 of 48 forums. Counting authors who name an adoption retrieves the reference in 17. Their paired outcomes, shown in Table I, imply a difference of 17 families, or 35.4 percentage points. The exact two-sided sign test on the 27 discordant families gives $p = 0.0015$. Conditional on a family entering this sample there is one paired observation per forum. Permuting 288 authors separately would answer a different question and would give far too little weight to agreement among relatives.

TABLE I
Agreement of two degree rules with independent nominations. An abstention counts as an incorrect retrieval.

Departure rule	Adoption correct	Adoption incorrect	Total
Correct	12	22	34
Incorrect	5	9	14
Total	17	31	48

The unsigned mention rule returns the reference in 27 groups. Fig. 1 shows both the coverage of the rules and the number of correct results. The host baseline takes the cook maintaining the original forum when that person is on the roster. A host outside the roster gives an abstention. We include the baseline because access to a large preexisting collection can make its keeper unusually visible in a corpus of recipes [16]. Host status was collected before the supplementary shelf opened.

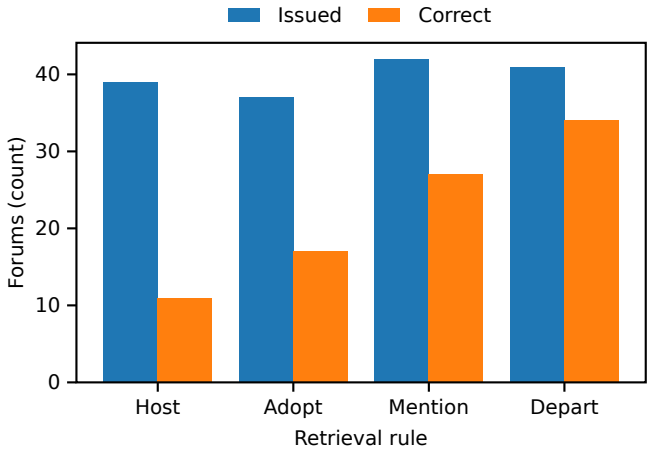


Fig. 1. Unique predictions and correct predictions, counted over 48 forums. Correct predictions are a subset of issued predictions.

Among the 34 correct departure predictions, ten reference cooks have no incoming adoption edge. An archive that required an accepted procedure before it would create an index entry would have no record of their authority. In six of these ten groups the reference supplied no card. Even recovering the text under all six contributor names would leave an empty result for the target's own procedure. It is the search for an exception to the procedure that returns a name.

Table II separates the two additional candidates from contributors. The denominator in the last row consists of families whose reference is one of the two people who did not supply a card. It does not count every absent cook in the sample. A contributor can receive at most five other contributors' edges while a noncontributor can receive six, a modest exposure advantage that motivated the sensitivity calculation below.

TABLE II
Correct top predictions by reference cook's relation to the archive. Each column counts families.

Reference cook	Families	Adopt	Mention	Depart
Contributed a card	30	14	18	20
Supplied no card	18	3	9	14
All	48	17	27	34

Dividing each degree by its number of eligible authors changes the departure rule in three forums. In two, an additional cook who had led outright becomes tied with a contributor. In the third a tie breaks in favor of a contributor. Across these changes one correct decision is lost and one gained. Equality of total scores conceals the altered identities, so an archive implementing such normalization should preserve both the numerator and the eligibility count. Fig. 2 presents the raw fractions behind the contrast in Table II.

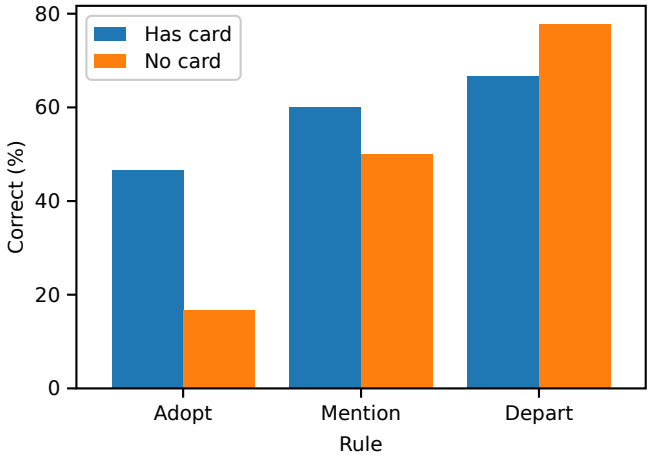


Fig. 2. Fraction of families with a correct retrieval, conditional on whether the reference supplied a card. Denominators are 30 and 18.

A. Sparse shelves

A researcher obtaining an archive after only three entries have arrived faces a different retrieval problem. We retain the original roster and nominations, sample subsets of the six cards and rerun each retrieval rule including the uniqueness check. We enumerate all subsets of a given size within each

family. Eq. (2) defines the recovery curve for rule R and sample size k. The units are families even though some sizes admit as many as twenty distinct shelves per family.

$$H_R(k) = \frac{1}{48}\sum_{g=1}^{48} \binom{6}{k}^{-1} \sum_{S \subset U_g:\,|S|=k} \mathbf{1}\{R(S) = v_g\}. \tag{2}$$

Fig. 3 gives the results of Eq. (2). The recovery probabilities for single cards are low in part because naming two relatives can generate a tie. A family made no more likely to agree by a fourth card can still become easier to index. The additional card may name a departure that was already present on the shelf, raising the corresponding degree above that of the competing cook. Once a full shelf has two equally supported candidates the abstention is final for purposes of our scoring.

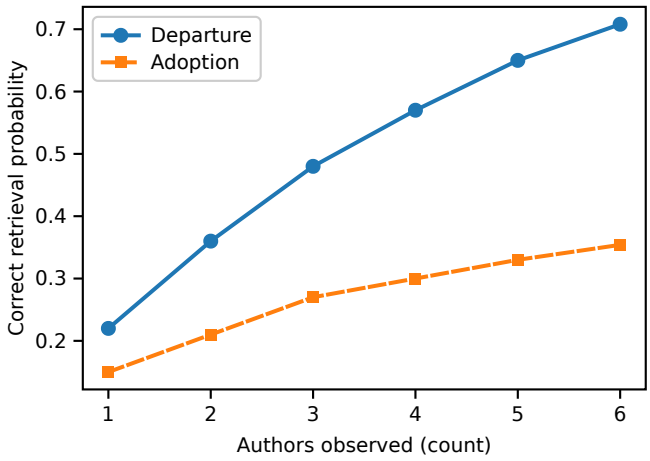


Fig. 3. Retrieval probability averaged over equal-sized subsets of authors in each family. Lines join discrete sizes; all families have equal weight.

VI. Reading the discrepant cases

We returned to the spans for the five families in which adoption alone retrieved the majority choice. In three, there was a recently circulated card with an explicit name in the title and a preparation credited in the text. The contributors could give an adopted operation with very little reconstruction. The counterclaim, when it appeared at all, generally concerned the way the dish was made somewhere outside the family. It had no named internal target. In the remaining two the reference cook had written the forum’s earliest instructions and routinely fielded questions about them. The other relatives gave the name as one gives the identifier of a document [17].

Among the 22 groups where departure alone succeeded, a recurrent difficulty was an apparently positive inheritance statement in the absence of any inherited operation. “Everything else is hers” follows a list of changes on several cards. We require a named operation for an adoption edge, so the residual reference receives no additional weight. Counting such blanket credits in a supplementary recoding raised the number of correct adoption predictions to 20. It left the identities of the three new successes quite clear, since the residual pronoun had a single antecedent in each case. It did little for a candidate referred to only in a prohibition.

In one group the reference’s sole documented instruction concerns the size at which the potatoes must be cut. Four writers describe a different size, each specifying the same comparison in a different unit. The record contains a thumbnail, the end of a teaspoon, a metric length and a comparison with the potatoes at another relative’s house. The last span supplies a second candidate, but no evidence about which end of the comparison is larger. A numerical reconstruction would have to drop the fourth account. The directed departure can be coded without resolving the units. Conversely, winning this ranking gives us no estimate of the actual dimensions of the reference potato.

We made an additional pass over all reference cooks who had no card of their own. In more than one collection the fieldworker asked whether the named person could have submitted a card. Relatives interpreted this as a request for a surviving written recipe and offered to ask at another house. We excluded all material acquired after closure. A retrospectively discovered card might be valuable evidence about the source cook while changing the data available to the original retrieval task. An archive should retain its deposit date separately from whatever date can be established for the paper [11].

VII. Interpretation and limits

Social information systems often use links to rank items for display. An editor viewing the family as a citation network might reasonably give a cook credit each time a recipe traces a specified operation to that cook. Our endpoint concerns a different search need. A relative could want to know which cook has to be mentioned in order for an unusual ingredient to count as a variation. A system tuned on confirmations will generally return some useful cards, but its errors are selective. A family member can be very well attested as someone who would never have included an ingredient, with nothing in the archive to show what else went into the bowl.

The majority nomination can reproduce a family hierarchy. A cook commonly named in contrast and a cook commonly consulted may both be consequences of age or perceived standing. The coding experiment identifies no effect of publicly departing from a recipe on future authority. Nominations preceded reading the shelf, and the same adults had many years to hear each other’s accounts before they completed their ballots. An intervention on the display of credits would require an outcome after the display and some way of distinguishing a change in search behavior from a change in esteem [18].

We excluded diffuse families by requiring a majority nomination, used one dish, and confined the roster to eight positions. Forcing a single target in a collection with two traditions can make the larger branch appear to define the smaller one. Branch labels would require several more respondents per group and could disclose relationships the owner chose to keep outside the shelf. Procedures for consent and aggregate reporting in small, identifiable networks differ from those used for

harvesting links between public pages [19], [20]. Our consent covered the group-local network and paraphrases of operations. We do not release named edges or a family-level ranking.

VIII. Archival implications

A negative attribution can identify a cook and preserve an instruction on which nobody on the shelf acts. The editor needs a field for the direction of departure and the person from whom the contributor departed. The existing source field is still needed for instructions a contributor actually used. The larger practical issue is the name authority file. Its custodian will occasionally have to create a permanent entry for a person for whom the archive can supply six corrections and no recipe.

Appendix A. Adjudication examples

The readers split an account at a change of attributed person or operation. If a clause carried a comparison of one relative’s version with another’s but did not locate the contributor’s own version, readers retained the span for interpretation and entered no outgoing edge. For instance, two relatives can disagree about leaving skins on with no evidence as to what the writer served. Inferring an outgoing tie from the relative mentioned last would reward a choice of word order. A statement that the writer uses the first relative’s method locates the writer and changes the coding.

An imperative inside indirect speech requires similar care. If a contributor reports that a parent has always told the family never to peel a potato, the sentence documents a named prohibition. It gives no sign by itself. A subsequent report that the writer peels them supplies a departure. A report that nobody at the table would touch them peeled can refer to anticipated guests rather than the author’s own dish. The coders did not infer the state of the potatoes from the behavior expected of those guests. The test for these cases is whether the card could truthfully accompany a dish made using either procedure.

One risk in a small graph is treating a group appellation as an individual. Where two names could correspond to “Gran”, the reader consulted a dictionary prepared at intake and specific to that writer. If the dictionary still admitted both candidates, the mention remained unresolved. Allocating a half edge to each would create a new kind of weighted tie. Nearly all ambiguities could have been resolved by asking the family later. Such a query might also have told the family which cook the researchers were trying to identify [21].

Appendix B. Uncertainty and ties

For the paired test in Section V we condition on 27 disagreements and use the equal-probability sign distribution under direction exchangeability across independent groups. Twice the probability of five or fewer successes in 27 trials gives the stated two-sided result. The test compares the two fixed rules. It does not compare an optimized mixture of their scores with either individual score. Trying all possible mixtures and then using the same sign distribution would invalidate that calculation.

With a blank included, an electorate of three has several routes to a missing majority. In particular a single name and two blanks leave no reference for evaluation. The retrieval algorithm may nevertheless issue a prediction. Reporting agreement on all 60 shelves would require a definition of correctness on those twelve and a new numerator. A rule that abstains often should not receive credit simply because the evaluators also found it difficult to agree on a name. The family with the empty result file retains its shelf and can still use it to look up an ingredient.

Acknowledgments

We thank the forum custodians who maintained the separate ballot file and checked the publication paraphrases.

References

[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[2] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[3] F. Heider, “Attitudes and Cognitive Organization,” *Journal of Psychology*, 1946.

[4] D. Cartwright and F. Harary, “Structural Balance: A Generalization of Heider's Theory,” *Psychological Review*, 1956.

[5] L. C. Freeman, “A Set of Measures of Centrality Based on Betweenness,” *Sociometry*, 1977.

[6] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[7] E. O. Laumann, P. V. Marsden, and D. Prensky, “The Boundary Specification Problem in Network Analysis,” in *Proc. Applied Network Analysis: A Methodological Introduction*, 1983.

[8] M. Sollinder, “Reference vertices outside a collection's membership roll,” *Journal of Relational Archives*, 2018.

[9] U. Raletto, “Departures recorded in household pattern books,” *Annals of Domestic Collections*, 2016.

[10] E. Qevani and I. Tomaret, “Naming the previous operator in craft notes,” *Studies in Archival Relations*, 2019.

[11] G. Mevrini, *Date of deposit in a retrospective family archive*. Low Shore Press, 2015.

[12] R. Dalverne, “Chronology and recollection in privately held recipe books,” *Annals of Domestic Collections*, 2020.

[13] A. Cazerin, “Local scope in annotated omission clauses,” *Studies in Archival Relations*, 2021.

[14] L. Saavare, “Resolving kin terms on a contributor-specific roster,” *Journal of Relational Archives*, 2022.

[15] A. Ferriton, “Agreement on rare edges in small annotated networks,” *Studies in Archival Relations*, 2017.

[16] L. Nastenek, “Host effects in the retrieval of a household collection,” *Journal of Relational Archives*, 2019.

[17] O. Ravellin, “People used as document identifiers in informal libraries,” *Annals of Domestic Collections*, 2022.

[18] I. Harovent, “Reference display and delayed choices of source,” *Journal of Relational Archives*, 2024.

[19] A. Zavrine, “Network extracts from invitation-only collections,” *Studies in Archival Relations*, 2022.

[20] P. Velmaris, *Publishing partial rosters of small domestic archives*. Low Shore Press, 2021.

[21] M. Tarvenne, “Researcher queries and retrospective authority records,” *Annals of Domestic Collections*, 2023.