Peeking Through the Homelessness System with a Network Science Lens

Charalampos Chelmis
Department of Computer Science
University at Albany, SUNY
Albany, NY, USA
cchelmis@albany.edu

Khandker Sadia Rahman
Department of Computer Science
University at Albany, SUNY
Albany, NY, USA
srahman2@albany.edu

***Abstract*—This paper models, for the first time, the homelessness system as a network of interconnected services which individuals traverse over time towards securing stable housing, and formalizes the concept of stability upon exit of the system. A computational analysis of individual–level longitudinal homelessness data shows that the ultimate goal is either reached quickly or not at all, regardless of starting conditions, indicating the importance of addressing the homeless' needs early on.**

***Index Terms*—Applied network science, computational social science, human services, socially important data science**

I. Introduction

Homelessness represents a long–standing, societal–scale problem with considerable individual and social costs [4]. While several studies have focused on the causes and contributors to homelessness [2], [9], only few have examined factors (e.g., length of stay) associated with outcomes (e.g., time–to–exit [8], likelihood of readmission [16], and housing stability [13]) once a homeless person has been admitted to the homelessness system. Unlike prior work, we assert that administrative data collected by homeless service providers can be represented as a rich network of interconnected homeless services which individuals traverse with the goal of securing stable housing. This unique view allows us to computationally analyze the efficiency and effectiveness of the homelessness system using network science principles. Our key contributions can be summarized as follows:

- We model the homeless system as a network of interconnected services which individuals traverse over time.
- We operationalize the concept of stability upon exit of the system with three alternative coding schemes.
- We computationally analyze a comprehensive set of individual–level longitudinal homelessness data from the Capital Region of the state of New York.
- We identify patterns that can explain the observed pathways of individuals through the homelessness system.
- For reproducibility, we make our source code available at https://github.com/IDIASLab/Homelessness.

II. Related Work

Prior work on factors that can cause or contribute to homelessness (e.g., [2], [6], [9], [11]), has mainly focused on structural determinants, with only few studies examining individual factors (e.g., [3], [6], [10]), mainly due to lack of individual–level data. On the other hand, predictors of outcomes (e.g., time–to–exit [5], [8] or likelihood of readmission [5], [16]) for a homeless person, once she has been admitted to the homelessness system, have been of interest. Similarly, variables (e.g., number of times being homeless in the past three years) indicative of one's potential for housing stability [7], [13], [15] at the time she is about to exit the homelessness system have also been explored. Our study builds upon this broad theme of work by operationalizing the concept of stable exit, providing the context (i.e., underlying network structure) in which factors are to be examined, and computationally analyzing the observed pathways of individuals through the homelessness system over time.

III. Data

The data we used for this study comprises an anonymized set of $50,469$ records of all services provided by homelessness service providers in the Capital Region of the state of New York from 2012 to 2018 to a total of $38,954$ individuals. The data was provided by CARES of NY Inc., a non–profit organization that locally manages the Homeless Management Information System[1] (HMIS), and captures (i) a timeline of services, including the beginning and end dates of each service, transitions between service types, exits and reentries (i.e., receipt of services after exiting the system), (ii) housing outcomes for those who exit the homeless services system, and, for those who were served again later by continuum

[1]HMIS is an information technology system founded by the U.S. Department of Housing and Urban Development (HUD).

This material is based upon work supported by the National Science Foundation under Grant No. ECCS–1737443.

ASONAM '21, November 8-11, 2021, Virtual Event, Netherlands

ACM ISBN 978-1-4503-9128-3/21/11...$15.00
http://dx.doi.org/10.1145/3487351.3488321

projects, lengths of time between exit and return to homelessness, (iii) demographic characteristics, such as age, race, gender, and veteran status, as well as household relations (e.g., household head or child), education history (e.g., last grade completed), and disabling condition (if any), and (iv) time–variant properties, such as monthly income and receipt of public assistance, and health information (e.g., mental state) [14]. Individuals are linked across services over time by a unique and anonymous identifier.

We focus on those individuals requesting services after exiting the system, i.e., *reentering* multiple times (15.4% of the total). We perform a longitudinal analysis to offer insights into the long–term progress (or lack thereof) of individuals towards stable housing. After addressing issues with erroneous and missing data, we were left with $18,818$ records of $6,011$ individuals that received services from 125 *project ids*[2] categorized in 9 *project types* [14].

IV. Modeling the Homelessness System as a Network

To better understand how individuals progress through the homelessness system over time, we assert that the embeddedness and interconnections of services must be taken into account. Specifically, we set forth to construct two transition graphs from the trails of completed *homelessness episodes*[3], namely, (i) a small network of project types, $\mathcal{G}_P$ and (ii) a larger network of project ids, $\mathcal{G}_I$. In both graphs, a directed edge from node i to j (where a node is either project type or id respectively) encodes at least one transition from i to j in the dataset. In total, 170 projects, uniquely identified by their id, are included in our dataset, and are grouped into 9 broader categories (i.e., *project types*). The construction of both graphs is nearly identical, as discussed below, with the only difference being nodes representing either project types or project ids, accordingly. Both graphs are directed and weighted, and may contain self loops, indicating a transition to the same program.

The construction of either transition graph starts with ordering enrollment records in acceding chronological order by entry date per individual, forming a *trajectory*, i.e., a sequence of enrollment records, each associated with a project id, and a project type, and corresponding entry and exit dates. We additionally use two special nodes, namely, entry and exit, to capture the state of individuals before entering and after exiting the homelessness system accordingly. Starting with the earliest entry date for a single individual, a new trajectory is created. This trajectory comprises the entry node, followed by the project type associated with the first enrollment. If this is the only enrollment for the given individual, node exit is added to the trajectory as the last transition, and the exit type (e.g., "Rental by client, no ongoing housing subsidy") is recorded. If more enrollments exist for the individual, the enrollment with the earliest entry date of the remaining enrollments is examined next. Thus, the first project type would be connected to the second one in the current trajectory. If an individual enters a new project prior to exiting from another, we record the transition by updating the exit date of the first enrollment. This process continues for every enrollment associated with an individual in order of entry date. The process is repeated until all episodes for every individual have been examined.

[2]A unique identifier is automatically generated by the HMIS at the time the project is created in the HMIS.

[3]Each episode is uniquely defined for each individual based on its corresponding entry and exit dates and service received.

In the second step, a transition graph is constructed using the list of trajectories computed in the first step. Specifically, there is one node for each of the nine project types, as well as one node for entry and one node for exit. A directed edge between nodes is created with a weight of one if there is at least one trajectory containing that specific transition. When a transition in a trajectory is encountered that has already been encoded as an edge in the graph, the weight of that edge is increased by one. After all trajectories have been examined and the graph construction is complete, edge weights are normalized so that all outgoing edges from a node sum to one (therefore serving as empirical transition probabilities).

Intuitively, the network of transitions among project types is more well connected than the fine–grain network of project ids. Specifically, the network of project ids is disconnected with 55 strongly connected components, the largest of which encompasses just $\sim 10\%$ of the nodes. The small diameter in both cases suggests that paths between those nodes that are connected are short. The existence of lengthy trajectories (c.f. Section VI) already hints towards inefficient paths that do not always progress forward towards stable housing.

TABLE I: High–level network statistics. Nodes "entry" and "exit" are not considered in this analysis.

Metric	$\mathcal{G}_P$	$\mathcal{G}_I$
Number of nodes	9	125
Number of edges	61	428
Density	0.847	0.028
Average degree	6.778	3.424
Min/Max in–degree	6/8	1/11
Min/Max out–degree	6/8	1/12
Diameter	2	4
Average path length	1.264	1.524
Number of WCCs	1	22
Size of largest WCC	100%	11.93%
Number of SCC	1	55
Size of largest SCC	100%	9.6%

V. Stable Exit Definition

A clear definition of *stable exit* is lacking in the literature. With the help of Dr. Wonhyung Lee, a *trained social scientist*, we explore three schemes based on individuals' destination after their exit, as shown in Table II. Table III provides a detailed description of exit destinations, and their categorization within each scheme. Specifically, Scheme 1 takes a binary approach, according to which destination types are categorized as either stable or unstable. Scheme 2 is based on HUD's destination types outlined in the HMIS [1] to characterize what is considered either positive or negative outcomes for street outreach, or permanent versus temporary destinations for all

TABLE II: Color coding of exit destinations by scheme. Hereafter, schemes are abbreviated as S_1, S_2, and S_3 respectively.

Scheme 1		Scheme 2		Scheme 3	
	Stable		Permanent		Ultimate goal
	Unstable		Temporary		Closer to exit
	Unknown		Institutional		Transitional phase
			Other		No progress or worse
					Hard to judge

TABLE III: Detailed overview of the proposed schemes (as defined in Table II).

ID	S_1	S_2	S_3	Exit destination description
1				Emergency shelter, including hotel or motel paid for with emergency shelter voucher, or RHY–funded Host Home shelter
2				Transitional housing for homeless persons (including homeless youth)
3				Permanent housing (other than RRH) for formerly homeless persons
4				Psychiatric hospital or other psychiatric facility
5				Substance abuse treatment facility or detox center
6				Hospital or other residential non–psychiatric medical facility
7				Jail, prison or juvenile detention facility
8				Clients does not know
9				Client refused
10				Rental by client, no ongoing housing subsidy
11				Owned by client, no ongoing housing subsidy
12				Staying or living with family, temporary tenure (e.g., room, apartment, or house)
13				Staying or living with friends, temporary tenure (e.g., room apartment or house)
14				Hotel or motel paid for without emergency shelter voucher
15				Foster care home or foster care group home
16				Place not meant for habitation (e.g., a vehicle, an abandoned building, us/train/subway station/airport or anywhere outside)
17				Other
18				Safe Haven
19				Rental by client, with VASH housing subsidy
20				Rental by client, with other ongoing housing
21				Owned by client, with ongoing housing subsidy
22				Staying or living with family, permanent tenure
23				Staying or living with friends, permanent tenure
24				Deceased
25				Long–term care facility or nursing home
26				Moved from one HOPWA funded project to HOPWA Permanent Housing
27				Moved from one HOPWA funded project to HOPWA Transitional Housing
28				Rental by client, with GPD TIP housing subsidy
29				Residential project or halfway house with no homeless criteria
30				No exit interview completed
31				Rental by client, with RRH or equivalent subsidy
99				Data not collected

other project types. Scheme 3 differentiates the degrees of progress, because not all temporary and institutional settings are the same. For instance, individuals in transitional housing can be defined as homeless because tenure in transitional housing is temporary and people must move at the end of the program, often, resulting in them exiting back into homelessness. In all cases, "unknown", "other", and "hard to judge" indicate that a conclusive determination cannot be made based on available information.

VI. Efficiency of the Homelessness System

The computed transition graphs (c.f. Section IV) provide an empirical estimate of the transition probabilities between project types (similarly ids). Fig. 1 shows how the distribution of the length of actual paths taken by the homeless (green) compares to *effective* (a.k.a. "always moving forward") trajectories (red), i.e., paths that ignore backward transitions (i.e., admission to a program type already admitted in the past) towards exiting the system. While the distribution of effective path lengths is tight around 2 steps, the empirical distribution of actual paths taken is heavy–tailed, with the majority of actual paths not being much longer than effective paths.

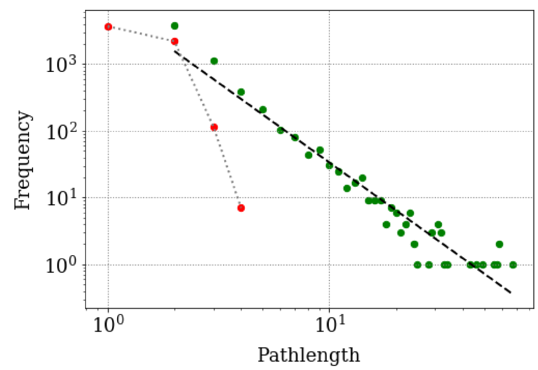

Fig. 1: Distribution of complete (green) and effective (red) trajectory lengths. Least square power–law fit with slope of -2.38 is shown in black.

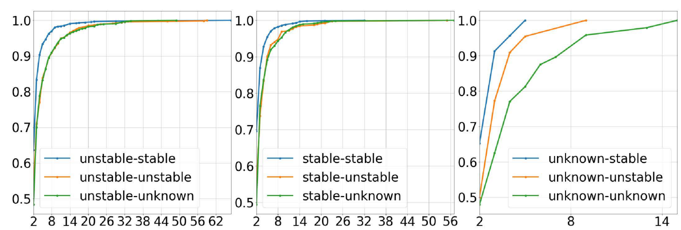

Fig. 2: Cumulative distribution function for path length as a function of entry and exit status, defined with respect to Scheme 1 presented in Tables II and III.

Since the same categorization is used in the HMIS for both the entry (i.e., before entering the homelessness system) and exit destination of individuals, we apply each scheme to better understand how individuals progress through the system based on their status upon entry. Fig. 2 shows the cumulative distribution of path lengths for those who enter the system from an unstable (left) or stable (center) situation, or unknown status (right) according to Scheme 1, to exit with a corresponding destination in k steps. In all cases, the distributions are heavy tailed, with unstable exits being reached when longer trajectories are taken, regardless of the starting point. Moreover, individuals starting from unstable conditions take longer trajectories before exiting the system to a stable destination. Finally, individuals exiting the system with an unknown destination follow longer paths on average when entering the system from stable or unknown conditions. We believe this to be indicative of individuals "falling through the cracks", i.e., not receiving the services or care they need to reach stable housing. According to Scheme 2, paths to permanent housing require relatively few transitions between project types in most cases, with the exception of temporary starting conditions. Similarly, an institutional exit destination is often reached quickly. After a more careful examination of the trajectories of those entering the system from a temporary starting point, it becomes clear that a temporary entry status leads to longer trajectories through the system than any other starting point. Finally, the path length of trajectories leading to transitional destinations in Scheme 3 are comparatively longer than hard to judge exit points. Similarly, those starting from a closer to exit position are better posed to reach the ultimate goal quickly. For those entering the system from a transitional phase or no progress or worse entry point, paths are longer overall. In most cases, paths leading to a no progress or worse exit are comparatively shorter than paths leading to other types of exit points, indicating that such individuals have potentially dropped out. The key insight is that the ultimate goal is either reached quickly (i.e., with few transitions) or not at all, regardless of starting conditions. In order for the homeless system to operate more efficiently and more effectively, the number of transitions an individual has to make within the homelessness system must be minimized.

A. Network Structure Versus Relevance

Next, we examine the trajectories of individuals exiting the system before reaching a stable exit. Fig. 3 shows the drop–out rate as a function of path position. Although presented for Scheme 3, we found the result to be consistent across all schemes. Intuitively, the probability of success is higher for shorter path lengths, whereas the probability of reentry increases drastically with the path position. At the same time, the probability of an individual "dropping out" of the system, slowly decreases with path length. This finding indicates both the perseverance of those remaining in the system, and the necessity to address the needs of the homeless early on.

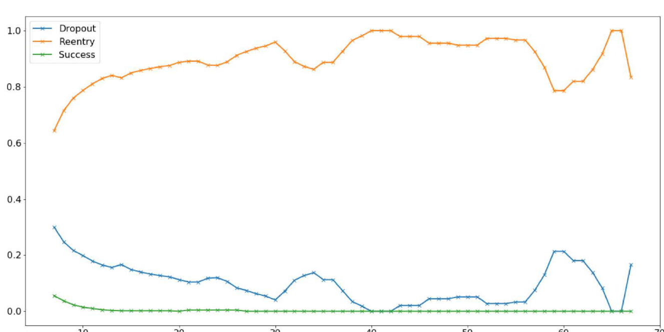

Fig. 3: Success, reentry, and drop–out probability ($y-$axis), as defined in Scheme 3, as a function of path position ($x-$axis).

Finally, we wish to understand whether the transition graph promotes (or hinders) trajectories towards achieving a positive outcome. Since no information in our dataset can be used to formalize a distance measure between project types (similarly for project ids), we look at trajectories as triples describing an individual, her assignment to a specific project type at a given time, and her feature vector. This view naturally lends itself to a heterogeneous network of three node types, namely individuals (I), projects (P), and features (F), with feature values treated as discrete. In order to obtain a representation of projects over the feature space, we traverse this tripartite network by computing the second power of the adjacency matrix (i.e., A^2), and approximate the similarity between two projects as the cosine of their TF–IDF vectors [12]. For each value of rank r, we consider all project pairs (i, j) such that j is the $r-$th most similar to i among all projects, and compute $P(r)$ as the fraction of these nodes for which i links to j.

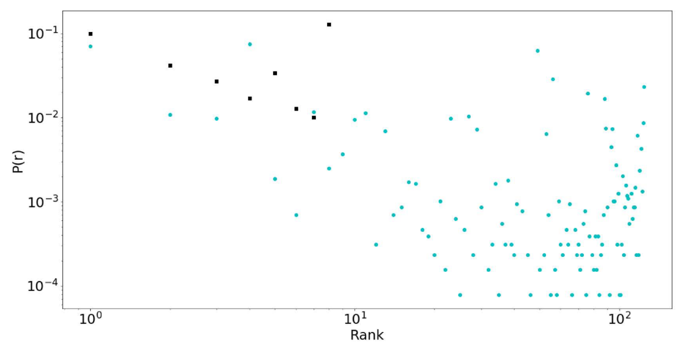

Fig. 4: Link probability, $P(r)$, as a function of rank, r, for the network of project types (black squares) and project ids (blue circles), accordingly.

Fig. 4 shows the link probability $P(r)$ for the networks of project types and ids accordingly, as a function of rank r. The overall trend in both cases suggests that the probability of a node (i.e., project types or ids) being connected with an outward edge to another node is higher the more related the two nodes are. This scenario is undesirable since "short–range" links between related project types (similarly for project ids) may not be particularly beneficial to the homeless in achieving their long–term goal of house stability. Specifically, transitioning to node j from i, with j being highly correlated to i (e.g., both being emergency shelters) would offer little help in achieving a positive outcome. On the other hand, the number of "long–range" edges between less related nodes, which would be potentially more helpful for getting closer to a stable exit, is limited. This leads us to conclude that the homeless services system is not conducive to efficient navigation because the majority of its edges emphasize connections across project types (similarly for ids) of similar function.

B. Progressing Towards Stable Exit

We conclude our analysis by taking a closer look at how the similarity between project types changes as individuals progress towards "ultimate goal", as defined in Scheme 3. Specifically, we compute from effective paths (i.e., ignore backward transitions) of the same length the cosine similarity

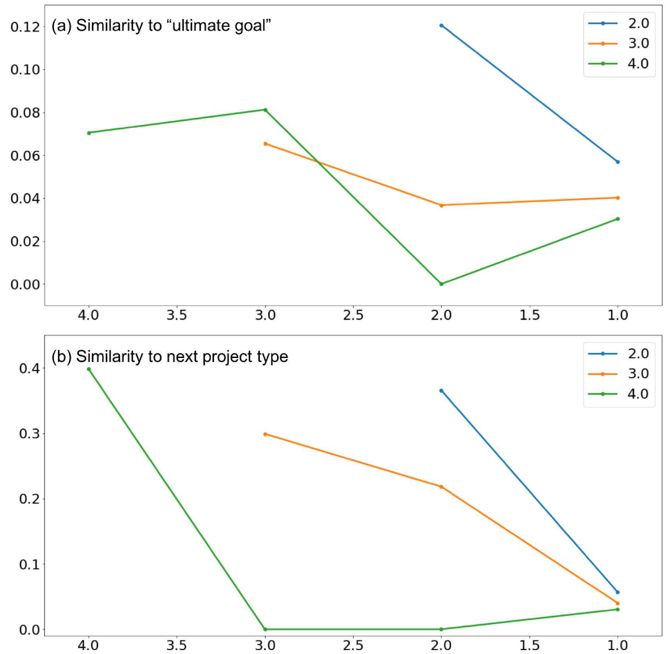

Fig. 5: Evolution of similarity as individuals progress towards “ultimate goal”. The x and y axes show respectively the distance to the target, and the cosine of the TF–IDF vectors.

of the TF–IDF vectors of the current project type and the target (Fig. 5a) or the following project type (Fig. 5b). The fact that the conceptual similarity (both between the past and current assignment, and with the target) decreases along paths provides additional evidence towards the need for “long–range” links to reach the goal of stable housing.

VII. Conclusion and Future Work

Contributions. In this paper, we used a one of a kind dataset of administrative records collected by homeless service providers to shed light into the progress of individuals once they enter the homelessness system towards securing stable housing. Apart from modeling the homelessness system as a network, which individuals traverse over time, we operationalized the notion of stable exit, which we subsequently used to analyze the efficiency of the network with respect to promoting trajectories leading to positive outcomes. We observed that, in its current form, the homelessness system is inefficient, with individuals reaching their goal quickly or not at all.

Limitations. Next, we would like to note the limitations of our present work that point to interesting future research directions. First, our observations are based on a single dataset geographically bounded to the Capital Region of the state of New York. Second, we rely on administrative data for receipt of services, which although indicate need, do not necessarily capture availability of services at any given time. Both these limitations may have resulted into a biased sample. Third, although we showed that the ability to achieve a positive outcome correlates with the network structure of the homelessness system, we do not claim to have fully revealed the causal effects of positive outcomes and other confounding variables, such as capacity constraints.

Future Directions. We hope, that by providing the foundation for reasoning about the homelessness system as a network, new functionality can be developed towards solving this thorny societal problem. We anticipate for instance further investigations in determining why individuals drop–out and developing algorithms to identify at risk individuals. Given the insights we offer here, another interesting direction is in designing recommendations methods for algorithmic homeless service delivery so as to unlock better outcomes.

Acknowledgment

We wish to thank Dr. Wonhyung Lee for assisting us in designing schemes to encode stable exit (c.f. Section V).

References

[1] “System Performance Measure 7: Destination Classification,” *Retrieved May 26, 2021, from https://www.hudexchange.info/resource/4966/system-performance-measure-7-destination-classification/*, 2016.

[2] M. Elliott and L. J. Krivo, “Structural determinants of homelessness in the United States,” *Social problems*, vol. 38, no. 1, pp. 113–131, 1991.

[3] S. Fitzpatrick, “Explaining homelessness: a critical realist perspective,” *Housing, Theory and Society*, vol. 22, no. 1, pp. 1–17, 2005.

[4] M. Henry, T. de Sousa, C. Roddey, S. Gayen, T. Joe Bednar, and A. Associates, “The 2020 Annual Homeless Assessment Report (AHAR) to Congress. Part 1: Point-in-time estimates of homelessness,” *The US Department of Housing and Urban Development*, 2021.

[5] B. Hong, A. Malik, J. Lundquist, I. Bellach, and C. E. Kontokosta, “Applications of machine learning methods to predict readmission and length-of-stay for homeless families: The case of win shelters in new york city,” *Journal of Technology in Human Services*, vol. 36, no. 1, pp. 89–104, 2018.

[6] J. Jarvis, “Individual determinants of homelessness: A descriptive approach,” *Journal of Housing Economics*, vol. 30, pp. 23–32, 2015.

[7] M. Johnstone, C. Parsell, J. Jetten, G. Dingle, and Z. Walter, “Breaking the cycle of homelessness: Housing stability and social support as predictors of long-term well-being,” *Housing Studies*, vol. 31, no. 4, pp. 410–426, 2016.

[8] R. Kuhn and D. P. Culhane, “Applying cluster analysis to test a typology of homelessness by pattern of shelter utilization: Results from the analysis of administrative data,” *American journal of community psychology*, vol. 26, no. 2, pp. 207–232, 1998.

[9] B. A. Lee, T. Price-Spratlen, and J. W. Kanan, “Determinants of homelessness in metropolitan areas,” *Journal of Urban Affairs*, vol. 25, no. 3, pp. 335–356, 2003.

[10] T. Main, “How to think about homelessness: Balancing structural and individual causes,” *Journal of social distress and the homeless*, vol. 7, no. 1, pp. 41–54, 1998.

[11] P. H. Rossi and J. D. Wright, “The determinants of homelessness,” *Health Affairs*, vol. 6, no. 1, pp. 19–32, 1987.

[12] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*. Cambridge University Press Cambridge, 2008, vol. 39.

[13] M. Shinn, B. C. Weitzman, D. Stojanovic, J. R. Knickman, L. Jimenez, L. Duchon, S. James, and D. H. Krantz, “Predictors of homelessness among families in New York City: from shelter request to housing stability,” *American Journal of Public Health*, vol. 88, no. 11, pp. 1651–1657, 1998.

[14] United States Department of Housing and Urban Development, “HMIS Data Standards Manual,” *Retrieved May 26, 2021, from https://www.hudexchange.info/resource/3824/hmis-data-dictionary/*, 2020.

[15] B. Van Straaten, J. Van der Laan, G. Rodenburg, S. N. Boersma, J. R. Wolf, and D. Van de Mheen, “Dutch homeless people 2.5 years after shelter admission: what are predictors of housing stability and housing satisfaction?” *Health & social care in the community*, vol. 25, no. 2, pp. 710–722, 2017.

[16] Y.-L. I. Wong, D. P. Culhane, and R. Kuhn, “Predictors of exit and reentry among family shelter users in New York City,” *Social Service Review*, vol. 71, no. 3, pp. 441–462, 1997.