S1. Detailed description of network measures and algorithms

Bipartite network

Data collected with VERITAS-Social can be represented as a bipartite graph $G = (V, E)$ with nodes separated in two disjoint and independent sets of nodes V_a and V_l and edges only connecting nodes of different sets. These sets contain the alters (V_a) and the visited locations (V_l), and the edges describe who are seen in each location. (Figure S1).

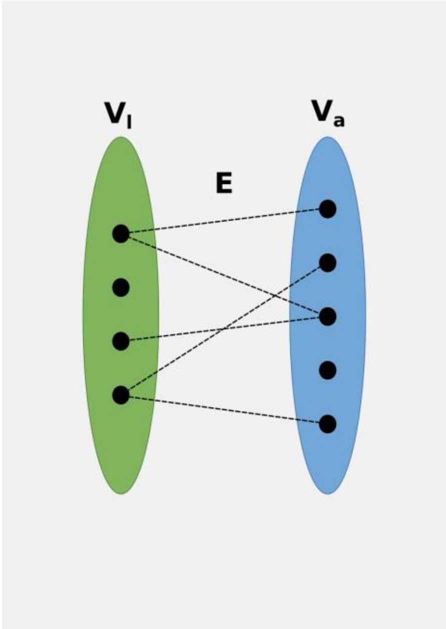

Figure S1: Visual example of a bipartite network composed of two sets of nodes V_a and V_l, and a set of edges E between the nodes of different sets. The two sets of nodes represent locations and alters, and the edges represent who are seen in each location. The nodes are positioned in two vertical layers to easily distinguish the two sets.

Size, degree and density

The *size* of the network $n = |V|$ refers to the number of nodes it contains. The number of alters and locations is denoted as $p = |V_a|$ and $q = |V_l|$. The number of edges in the networks is denoted by $m = |E|$. The *neighbourhood* N_i is the set of nodes directly connected to i. The *degree* of the node i is the number of its neighbours $k_i = |N_i|$. The average degree within V_a and V_l can be calculated as $\bar{k}(V_a) = {}^{m}/_{p}$ and $\bar{k}(V_l) = {}^{m}/_{q}$. The graph *density* $d = {}^{m}/_{pq}$ is the ratio between the number of edges and the maximum possible numbers of edges in the network.

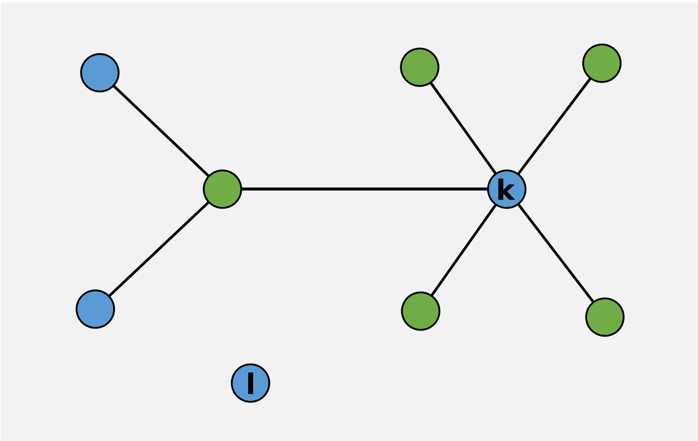

Figure S2: Visual example of size, degree and density in a hypothetical bipartite network of alters (blue nodes) and locations (green nodes). The total size of the network is 9, separated into 3 alters and 6 locations. The degree of node k is 5, which means that alter k is connected to 5 locations. Node l is disconnected from the rest of the network, which means that it has a degree of 0. There are 7 edges in the network, while the maximum possible total number is 27. Therefore, the network has a density of 7/27 = 0.26.

Community structure

Community structure is a concept that stems from the observation that real-world networks tend to be organized in groups of nodes with common properties. The problem is that this concept comes from an observation about regularities in empirical systems and there is no actual consensus on a formal definition. The most common understanding is that nodes are more likely to share edges within their own communities than with the rest of the network (Fortunato and Hric, 2016). But concretely, communities are mostly defined in an operational way, as they are the end product of community detection algorithms (Fortunato, 2010). In this paper, we used the LP&BRIM algorithm of Liu and Murata (2009), built on Barber's modularity (2007) and the label propagation algorithm (Raghavan et al., 2007), to identify non-overlapping communities composed of alters and locations. This algorithm tries to find the maximum value of the network *modularity* (Q), a structural estimate of the network partition quality. In short, the modularity measures the difference between the actual and the expected number of edges within communities (see Newman and Girvan (2004) for a detailed description of modularity, and Barber (2007) for its calculation in bipartite network). An initial partition is first computed using an adaptation of the label propagation algorithm. Each node of a set (V_a or V_l) is labelled with its own community, producing as many communities as there are nodes. Then, recursively, each node of the other set is labelled by the community most shared among its neighbours. The algorithm stops when each node is labelled with the community that is shared by most of its neighbors. The initial partition is then further refined using the bipartite recursive induced module algorithm (BRIM). It uses the spectral properties of the network to recursively relabel nodes of each set V_a and V_l, and it stops when a local maximum of Q is found (See Barber (2007) for a detailed description of this algorithm). For each network, the LP&BRIM algorithm was run 1000 times and the partition with the highest modularity value was retained.

Two additional indicators were calculated to assess the position of the nodes relatively to the identified communities (Guimerà and Amaral, 2005). *Within-module degree* is a z-score that the connectivity of a node with the other nodes of its community, defined as

$$z_i = \frac{k_i - \bar{k}_{s_i}}{\sigma_{s_i}}$$

where $\bar{k}_{s_i}$ and σ_{s_i} are the average and the standard deviation of degrees in the community s of the node i. Two nodes having the same z-score can have different positions in the whole network, since a node can also have connections with nodes of other communities. The *participation index* measures how nodes are located between the different communities in the network, defined as

$$p_i = 1 - \sum_{s=1}^{c} (\frac{k_{is}}{k_i})^2$$

where k_{is} is the number of neighbors of node i in community s. The participation index takes a value of 0 if all edges of a node are within the same community and tend towards 1 if all edges are evenly distributed across all communities. The maximum value of the participation index depends on the number of communities, and equal to $1 - \frac{1}{c}$. In participants A and B bipartite networks, the maximum value is 0.75 as both networks are divided into 4 communities. See (Guimerà and Amaral, 2005) for a detailed discussion on *within-module degree* and *participation index*.

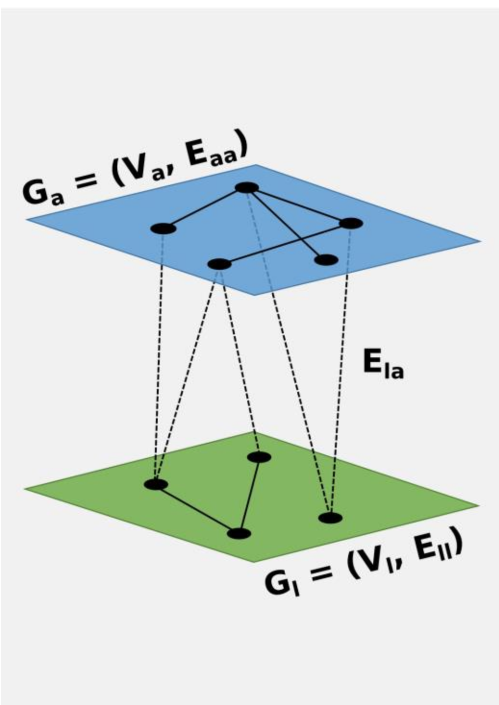

Figure S3: Visual representation of a system of interconnected networks composed of two sub-networks $G_l = (V_l, E_{ll})$ and $G_a = (V_a, E_{aa})$, representing a spatial and a social network, and a set of edges E_{la} between the sub-networks. Nodes are positioned in two horizontal layers to easily distinguished the two sub-networks.

interconnected networks

VERITAS-Social data can also be represented as interconnected networks – a social network $G_a = (V_a, E_{aa})$ and a spatial network $G_l = (V_l, E_{ll})$ connected by a set of edges E_{al}. The social network contains the alters V_a connected by interpersonal relationships E_{aa} (i.e., who know whom). The spatial network contains the visited locations V_l connected by spatial relationships E_{ll}. The latter was not defined in this paper. The connections between alters and visited locations E_{al} are defined by who is seen in each location (Figure S1).

Cross-clustering coefficient

The *cross-degree* is, for a node i from a sub-network, the number of its neighbours in the other sub-network. The cross degree for alters k_i^{al} and locations k_j^{la} are defined as follow

$$k_i^{al} = \left|N_i^{al}\right| : i \in V_a$$
$$k_j^{la} = \left|N_j^{la}\right| : j \in V_l$$

where N_i^{al} and N_j^{la} are the neighbours in the other sub-network. It is equivalent to the degree calculated in a bipartite network (their can be no connections between two neighbours of a same set).

The *cross-clustering coefficient* is the probability that two neighbours in the other sub-network are also connected to each other. Since edges between locations E_{ll} have not been defined in this paper, the cross-clustering coefficient has only been calculated for locations, defined as

$$cc_j^{la} = \frac{2\left|E(N_j^{la})\right|}{k_j^{la}\left(k_j^{la} - 1\right)}$$

where $E(N_j^{la})$ is the set of edges between neighbours of location j in the social network G_a (Donges et al., 2011). This cross clustering coefficient is equivalent to the clustering coefficient calculated in a unipartite network (Watts and Strogatz, 1998). See Figure S2 for a visual illustration of the cross-clustering coefficient.

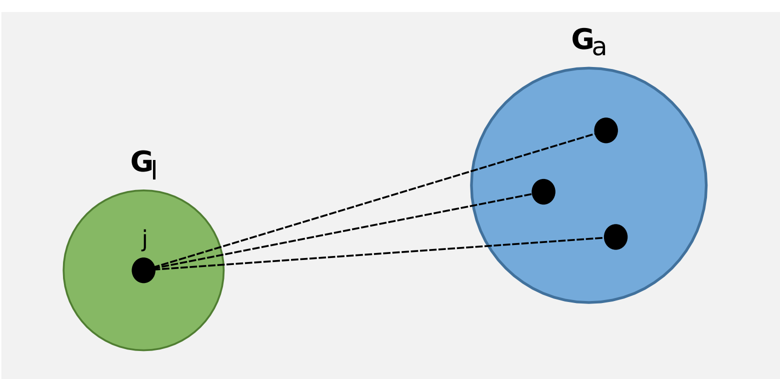

Figure S4: Visual illustration of the cross-clustering coefficient. For the node j within subnetwork G_l, $cc_j^{la} = \frac{2*1}{3(3-1)} = \frac{2}{6} = 0.33$

S2. Node-level network measures

	Bipartite				Interconnected
Node	Degree	Clusters	Within z-score	Participation index	Cross-clustering coefficient
A1	5		1.79	0.32	NA
A2	1		-0.87	0	NA
A3	3		0.87	0.44	NA
A4	1		NA	0	NA
A5	1		-0.33	0	NA
A6	1		-0.33	0	NA
A7	1		-0.33	0	NA
A8	1		-0.33	0	NA
A9	1		-0.33	0	NA
A10	1		-0.33	0	NA
A11	1		-0.33	0	NA
A12	1		-0.33	0	NA
L1	1		-0.45	0	NA
L2	1		-0.45	0	NA
L3	1		-0.45	0	NA
L4	1		-0.45	0	NA
L5	1		-0.45	0	NA
L6	3		0.87	0.44	0.67
L7	1		-0.87	0	NA
L8	1		*NA*	0	NA
L9	9		2.67	*0.2*	NA
L10	0		*NA*	*NA*	NA
L11	0		*NA*	*NA*	NA
L12	0		*NA*	*NA*	NA

Table S1: Node-level measures for participant A's bipartite and interconnected networks. Alters and locations are labelled Ax and Lx respectively.

	Bipartite					Interconnected
Node	Degree	Clustering Coefficient	Dependency	Within z-score	Participation index	Cross clustering coefficient
A13	2	0.42	0	NA	0	NA
A14	3	0.36	0	NA	0.44	NA
A15	1	0.58	0	-0.45	0	NA
A16	1	0.58	0	-0.45	0	NA
A17	1	0.58	0	-0.45	0	NA
A18	2	0.42	0	-0.45	0.5	NA
A19	11	0.14	5	2.04	0.68	NA
A20	2	0.47	0	-0.06	0	NA
A21	2	0.47	0	-0.06	0	NA
A22	1	0.69	0	-0.48	0	NA
A23	1	0.69	0	-0.48	0	NA
A24	1	0.69	0	-0.48	0	NA
A25	1	0.69	0	-0.48	0	NA
A26	1	0.69	0	-0.48	0	NA
A27	1	0.69	0	-0.48	0	NA
A28	2	0.47	0	-0.06	0	NA
A29	2	0.47	0	-0.06	0	NA
A30	0	NA	0	NA	NA	NA
A31	0	NA	0	NA	NA	NA
L13	3	0.34	0	NA	0.44	1
L14	3	0.34	0	NA	0.44	1
L15	6	0.18	3	1.79	0.5	0.4
L16	1	0.56	0	-0.41	0	NA
L17	1	0.56	0	-0.41	0	NA
L18	1	0.56	0	-0.41	0	NA
L19	1	0.56	0	-0.41	0	NA
L20	1	0.56	0	-0.41	0	NA
L21	3	0.26	0	-0.06	0.44	1
L22	3	0.26	0	-0.06	0.44	1
L23	12	0.12	6	3.26	0.29	1

Table S2: Node-level measures for participant B's bipartite and interconnected networks. Alters and locations are labelled Ax and Lx respectively.

S3. Alters and locations characteristics

Tables S3 to S6 present the characteristics of alters and locations reported by participants A and B. In the tables describing *alters* (Table S3 and S5), different variables are used to describe people and groups. For people, *age* in number of years, *gender* in binary format, *relation type* in 6 categories (i.e., spouse, child, other family member, friend, acquaintance, work colleague) and *interaction frequency* in number of times per year are provided. For *groups*, the *relation types* of most group members, *interaction frequency*, *group size* and *group description* in verbatim transcription are provided. For *locations* (Table S4 and S6), *type of location* in 25 categories used to collect activities during CURHA, *frequency of attendance* in number of times per year, *usual mode of transportation* in 4 categories (i.e., walking, car, car other, public transportation), *geographical location* and *distance from home* are provided. Geographic locations are defined according to the Sherbrooke boroughs and the surrounding administrative municipalities. These boundaries were taken from the SDA20K dataset available in the Quebec Open Data Portal (Données Québec, 2019). Distances from home were calculated as shortest routes according to the road network with GRASS 7.0.4 (Neteler et al., 2012) and using CanMap Streetfiles V2013.3 (DMTI Spatial Inc, 2013).

Node	Type	Age	Gender	Relation	Frequency of interactions	Group size	Group descriptive
A1	Person	55	F	Children	Daily	*NA*	*NA*
A2	Person	21	F	Family	Daily	*NA*	*NA*
A3	Person	84	F	Friend	Weekly	*NA*	*NA*
A4	Person	86	F	Friend	Weekly	*NA*	*NA*
A5-A6	Group	*NA*	*NA*	Friend	Weekly	4	Volunteering
A7-A12	Group	*NA*	*NA*	Friend	Weekly	34	Choral

Table S3: Characteristics of participant A's alters

Node	Type	Frequency	Transportation	Location	Distance
L1	General practitioner	Annually	Car other	Mont-Bellevue	13.08
L2	Shopping centre	Monthly	Car other	Jacques-Cartier	9.88
L3	Medical specialist	Semiannually	Car other	Fleurimont	11.74
L4	Restaurant	Monthly	Car other	Jacques-Cartier	11.27
L5	Grocery store	Weekly	Car other	Jacques-Cartier	9.74
L6	Home	*NA*	*NA*	Bromptonville	*NA*
L7	Alter residence	Monthly	Walking	Bromptonville	0.54
L8	Alter residence	Semimonthly	Walking	Bromptonville	1.06
L9	Church	Weekly	Walking	Bromptonville	0.59
L10	Pharmacy	Weekly	Walking	Bromptonville	1.1
L11	Grocery store	Weekly	Walking	Bromptonville	0.59
L12	Bank	Monthly	Walking	Bromptonville	0.46

Table S4: Characteristics of participant A's visited locations

Node	Type	Age	Gender	Relation	Frequency of interactions
A13	Person	50	F	Family	Weekly
A14	Person	50	M	Children	Weekly
A15	Person	75	M	Friend	Monthly
A16	Person	65	M	Acquaintance	Monthly
A17	Person	60	M	Friend	Monthly
A18	Person	60	M	Friend	Weekly
A19	Person	85	F	Spouse	Daily
A20	Person	75	M	Friend	Weekly
A21	Person	75	F	Friend	Weekly
A22	Person	65	F	Friend	Weekly
A23	Person	60	M	Friend	Weekly
A24	Person	60	F	Friend	Weekly
A25	Person	60	F	Friend	Weekly
A26	Person	60	F	Family	Weekly
A27	Person	70	M	Friend	Weekly
A28	Person	67	F	Friend	Weekly
A29	Person	70	M	Friend	Weekly
A30	Person	55	M	Friend	Monthly
A31	Person	70	M	Friend	Annually

Table S5: Characteristics of participant B's alters

Node	Description	Frequency	Transportation	Location	Distance
L13	Home	*NA*	*NA*	Newport	*NA*
L14	Visiting people	Monthly	Walking	Newport	0.54
L15	Restaurant	Semimonthly	Car other	Lennoxville	35.95
L16	Bank	Quarterly	Car other	Cookshire-Eaton	12.82
L17	General practitioner	Quarterly	Car other	East Angus	22.40
L18	Convenience store	Semimonthly	Car other	Cookshire-Eaton	4.84
L19	Convenience store	Semimonthly	Car other	Cookshire-Eaton	4.92
L20	Pharmacy	Monthly	Car other	Cookshire-Eaton	12.89
L21	Visiting people	5 times per year	Car other	Newport	11.39
L22	Visiting people	Bimonthly	Car other	Newport	7.70
L23	Church	Weekly	Car other	Cookshire-Eaton	4.77

Table S6: Characteristics of participant B's visited locations

S4. Including groups in network analysis

With VERITAS-Social, alters can either represent individually declared or groups of people. Including both forms of alters can be problematic during network analysis. Treating groups as individual nodes underestimates the number of people encountered in certain locations, whereas treating them as *n* nodes, where *n* is the size of the group, may have a too strong impact on the size of the network and its underlying properties. To illustrate, for respondent A, two groups, consisting of 34 and 4 people, are seen in location L9. If we consider each of these groups as a single node, L9 has a degree of 3 and there is 6 alters in the network. If we consider these groups as *n* nodes, the degree of L9 and the number of alters change to 39 and 44, respectively. These marked differences will have an impact on the properties of the network, such as density and community structure. We suggest that weighting the number nodes by first transforming *n* (e.g., $\sqrt{n}$) can provide a more appropriate estimate of social exposure across locations while reducing the effect of large groups (in the Canadian CURHA sample, the largest reported group is 150 individuals). Nevertheless, the reported individuals and group members may not play similar roles (e.g., group members not reported individually may be acquaintances providing few health-relevant social resources), although these relationships will be considered as equivalent in this analysis.

References

Barber, M.J., 2007. Modularity and community detection in bipartite networks. Physical Review E 76, 066102. https://doi.org/10.1103/PhysRevE.76.066102

DMTI Spatial Inc, 2013. CanMap Streetfiles. CanMap.

Donges, J.F., Schultz, H.C.H., Marwan, N., Zou, Y., Kurths, J., 2011. Investigating the topology of interacting networks. Eur. Phys. J. B 84, 635–651. https://doi.org/10.1140/epjb/e2011-10795-8

Données Québec, 2019. Quebec open data portal. URL https://www.donneesquebec.ca/ (accessed 1.15.19).

Fortunato, S., 2010. Community detection in graphs. Physics Reports 486, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Fortunato, S., Hric, D., 2016. Community detection in networks: A user guide. Physics Reports, Community detection in networks: A user guide 659, 1–44. https://doi.org/10.1016/j.physrep.2016.09.002

Guimerà, R., Amaral, L.A.N., 2005. Functional cartography of complex metabolic networks. Nature 433, 895. https://doi.org/10.1038/nature03288

Liu, X., Murata, T., 2009. Community Detection in Large-Scale Bipartite Networks, in: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT ’09. IEEE Computer Society, Washington, DC, USA, pp. 50–57. https://doi.org/10.1109/WI-IAT.2009.15

Neteler, M., Bowman, M.H., Landa, M., Metz, M., 2012. GRASS GIS: A multi-purpose open source GIS. Environmental Modelling & Software 31, 124–130. https://doi.org/10.1016/j.envsoft.2011.11.014

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical review E 69, 026113. https://doi.org/10.1103/PhysRevE.69.026113

Raghavan, U.N., Albert, R., Kumara, S., 2007. Near linear time algorithm to detect community structures in large-scale networks. Physical review E 76, 036106. https://doi.org/10.1103/PhysRevE.76.036106

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’networks. nature 393, 440–442. https://doi.org/10.1038/30918