

SOFTWARE

Open Access

GatewayNet: a form of sequential rule mining



Phillip C. S. R. Kilgore^{1*†} , Nadejda Korneeva^{2,3}, Thomas C. Arnold³, Marjan Trutschl^{1,4†} and Urška Cvek^{1,4†}

Abstract

Background: The gateway hypothesis (and particularly the prediction of developmental stages in drug abuse) has been a subject of protracted debate since the 1970s. Extensive research has gone into this subject, but has yielded contradictory findings. We propose an algorithm for detecting both association and causation relationships given a discrete sequence of events, which we believe will be useful in addressing the validity of the gateway hypothesis. To assess the gateway hypothesis, we developed the GatewayNet algorithm, a refinement of sequential rule mining called initiation rule mining. After a brief mathematical definition, we describe how to perform initiation rule mining and how to infer causal relationships from its rules (“gateway rules”).

We tested GatewayNet against data for which relationships were known. After constructing a transaction database using a first-order Markov chain, we mined it to produce a gateway network. We then discuss various incarnations of the gateway network.

We then evaluated the performance of GatewayNet on urine drug screening data collected from the emergency department at LSU Health Sciences Center in Shreveport. A de-identified database of urine drug screenings ordered by the department between August 1998 and June 2011 was collected and then restricted to patients having at least one screening succeeding their first positive drug screening result.

Results: In the synthetic data, a chain of gateway rules was found in the network which demonstrated causation. We did not find any evidence of gateway rules in the empirical data, but we were able to isolate two documented transitions into benzodiazepine use.

Conclusions: We conclude that GatewayNet may show promise not only for substance use data, but other data involving sequences of events. We also express future goals for GatewayNet, including optimizing it for speed.

Keywords: Initiation rules, Gateway hypothesis, Association rule mining, Causal network, Structure learning

Background

The *Gateway hypothesis* (also *gateway theory* or *stepping-stone theory*) is the assertion that the use of certain psychoactive drugs (e.g., tobacco, alcohol, or cannabis) increases the likelihood that other drugs will later be used. It is commonly interpreted to mean that usage of one drug will encourage the initiation (or first usage) of new substances, and the first drug is therefore said to be called

a *gateway drug*. Another prediction that has been associated with the gateway hypothesis is that initiation for specific drugs develops in stages.

The gateway hypothesis

This hypothesis is controversial amongst substance abuse experts, as many studies with conflicting results have been released since intense interest beginning in the 1970s. For instance, Kandel originally predicted a chain of drug use progression from tobacco and alcohol to cannabis, then to LSD, amphetamines, or heroin. She posits that this association is bidirectional and that a similar sequence will occur for regression in drug use [1]. In 1984, a follow-up was performed to address the fact that detailed monitoring of adolescents into young adulthood, suggesting that

*Correspondence: pkilgore@lsus.edu

†Phillip C. S. R. Kilgore, Marjan Trutschl and Urška Cvek contributed equally to this work.

¹Department of Computer Science, LSU Shreveport, 1 University Place, 71115 Shreveport, USA

Full list of author information is available at the end of the article



initiation risk may be partially conditional on age and that risk progresses in stages [2]. At the height of the crack cocaine epidemic, Kandel and Yamaguchi reformed their model to account for its sudden appearance and found that a) cocaine precedes crack cocaine, and b) models using cocaine or crack cocaine exclusively had a poorer fit than those containing both [3].

O'Donnell and Clayton directly claimed a causal connection between marijuana and heroin use [4]. To support this, they note that marijuana and heroin are statistically associated, that marijuana precedes heroin use, and that this association is not spurious. O'Donnell and Clayton alleged that a large cohort of sociologists were skeptical of the gateway hypothesis at the time, and they argued that marijuana causes heroin use according to how sociologists understand causation [4].

Early criticism of causal predictions of the gateway hypothesis takes two major forms: that the evidence does not support the assertion or that the assertion is structurally flawed. In an attempt to replicate Kandel's work, Baumrind obtained a different pathway which implicated that tobacco succeeded cannabis (though both found that the use of socially-accepted substances precedes that of the unacceptable), noting that drug initiation order may be influenced by sociocultural aspects [5].

An additional form of criticism arose in the way that the conclusion itself was being formulated. In an article warning against drawing false conclusions of causation, Baumrind cites O'Donnell and Clayton as an exemplar of this [6]; she later comments that Guttman scales cannot be extrapolated into a sequence of development stages as was done in Kandel's work [5]. Vanyukov et al. argue that the gateway hypothesis may lack falsifiability and that the concept itself is vague [7].

Nonetheless, contemporary support of the gateway hypothesis is mixed. It is known that rats exposed to $\Delta 9$ -tetrahydrocannabinol (THC, the primary active compound in cannabis) will increase self-administration of nicotine, heroin, and morphine [8–10], showing that cannabis can operate as a gateway drug outside of any particular cultural context. Conversely, it has been argued that the apparent progression is one of several, and that common liability to addiction may be enough to explain patterns in substance use [7]. One longitudinal study of New Zealand children concluded that although there was strong association with a diverse use of other drugs and that this may support a causal model, the underlying causal mechanisms are not well understood [11].

There are two major approaches involving longitudinal data used to assess drug use in human subjects: through self-reporting and through urine drug screening (UDS). In self-reporting studies, subjects are asked to inform investigators about their drug history. This method frequently tracks subjects from adolescence into adulthood

to determine both trends in usage and initiation. However, it may be influenced by response bias common to interviews and surveys [12, 13].

UDS detects metabolites associated with certain drugs use (usually via a panel assay). This offers a major advantage over self-reporting: it is possible to collect information that would otherwise be withheld in a self-reporting study. It also becomes possible to collect data from subjects who are unable to participate in interviews, such as infants (who are unlikely to consciously participate in drug use, but which may reveal drug use by parents).

The main disadvantage of this method is false positive results arising from misidentification of metabolites in urine. For instance, it is known that quinolone antimicrobials can create false positives for opiate presence [14, 15]. Several forms of medication (both prescribed and over-the-counter) are known to trigger false positives in drug tests; ibuprofen, a common analgesic, may trigger false positives for phencyclidine (PCP), cannabinoids, and barbiturates in some screening panels [15].

Previous approaches

The goal of GatewayNet is to predict initiation events and select those relationships which may be causal; therefore, it is important to consider past approaches to this problem. It should be noted that the causation referred to here is not deterministic causation: observation does not support the idea that a gateway drug is always followed by its target. Instead, the idea of *probabilistic causation* (i.e., event a is likely to cause b) is considered [16].

$$p(b|do(a)) > p(b|do(\neg a)) \quad (1)$$

Statistical treatment of this problem has been attempted in the literature. A simple method uses a linear probability model [11, 17], such as the one suggested by Beenstock and Rahav to predict how cigarettes influenced cannabis use Eq. 2, where S_{nt} and C_{nt} are indicators of cigarettes and cannabis respectively by sample n at time t , X is a vector of personality characteristics, D_y is the birth cohort for year y , and u_{nt} accounts for unobserved error. The gateway hypothesis predicts that if C is a gateway into S , then $\beta > 0$ [17].

$$S_{nt} = \alpha X_{nt} + \beta C_{n(t-1)} + \gamma_y D_y + u_{nt} \quad (2)$$

Hazard analysis has also been used to assess this problem [17]. In relation to the gateway hypothesis, hazard analysis attempts to ascertain the risk of initiating the use of another drug. Recently, latent transition analysis has been used to assess gateway relationships [18].

Bayesian inference is often used to assess claims of causation. For instance, a Bayesian method was applied to assess data from Norwegian young adults and yielded the

conclusion that proneness and accessibility are important contributing factors to hard drug use [19]. Another potential avenue might be in the form of a Bayesian Belief Network (BBN), a directed acyclic graph describing the probability of condition b occurring given condition a [20]; however, the literature does not record such an application of BBNs to the gateway hypothesis.

Association rule mining

Association Rule Mining (ARM) is a well-known method where a set of items called a *transaction* can be mined to produce *association rules* of the form $a \rightarrow b$, which is a prediction that when a is present, b will co-occur. A related strategy, known as *sequential rule mining* (SRM), can be used to predict that a will *precede* b in sequence. Algorithms which use SRM include the Co-occurrence Maps with Sequence PAttern Mining using Equivalent class (CM-SPADE) [21], Sequential PAttern Mining (SPAM) [22], and Closed Sequential Patterns (ClaSP) [23] algorithms.

Sequential rule mining is applicable to a problem such as the Gateway Hypothesis because the latter predicts a causal relationship; if a causes b , then it is necessary for a to precede b . Causation also implies that the first instance of b will not precede the first instance of a .

We claim three contributions to the literature: i) the application of sequential rule mining to the assessment of the Gateway hypothesis, ii) the use of these rules to construct a *gateway network* describing interaction between, and iii) the introduction of the *certainty* measure.

Implementation

To better understand the extent to which the Gateway Hypothesis manifests itself in drug use trends, we developed GatewayNet, an algorithm that constructs a directed, weighted graph of drug initiation events derived from a form of association rule mining. We then performed an evaluation against two data sets: a synthetic data set, and an empirical data set derived from UDS data.

Mathematical model

In the following paragraphs, the mathematical basis for GatewayNet (and in particular, initiation rule mining) are described. How this model is defined is critical to interpreting GatewayNet’s results, so it is described in detail here.

Precedence Relations Let E denote a set of events, $S : t \in \mathbb{Z}^+ \mapsto E$ denote a sequence of events called the *history* such that $S_t \subseteq E$ is the set of events occurring at some time t , $a \subseteq E$, and $b \subseteq e$. The predicate $a < b$ means “ a precedes b ” and is defined in Eq. 3.

$$a < b \equiv \exists t \in \mathbb{Z}^+ : a \subseteq S_t \wedge b \subseteq S_{t+1} \quad (3)$$

It should be noted that $a < a$ may yield true under this definition. The operand a is called the antecedent, while b is called the subsequent.

Initiation Relations Let $S_b^a = S_a \cup \dots \cup S_b$. For brevity, $S_t^* = S_t^1$ and $S^* = S_{|S|}^*$. The predicate $a \subseteq b$ means “ a initiates b ” (an instance thereof being called an *initiation rule*) and is defined in Eq. 4. An initiation rule $a \subseteq b$ has a degree which is the maximum between the number of elements in a and the number of elements in b Eq. 5.

$$a \rightarrow b \equiv \exists t : a < b \wedge b \not\subseteq S_{nt} \quad (4)$$

$$\equiv \bigvee_{t=1}^{|S|-1} a \subseteq S_t^* \wedge b \subseteq S_{t+1} \in b \not\subseteq S_t^*$$

$$\text{deg}(a \rightarrow b) \equiv \max(|a|, |b|) \quad (5)$$

Note that (unlike precedence relations) the initiation relation $a \rightarrow a$ is universally false. This relation can be further generalize d into windowed initiation. Let $z \in \mathbb{Z}^+$, $z' = z - 1$, and $a \xrightarrow{z} b$ denote an initiation rule within window z . In this generalization, only the most recent z time points are searched for the antecedent in every time-step. Because $a \xrightarrow{0} b$ is trivially false according to Eq. 6, it has been redefined Eq. 7.

$$a \xrightarrow{z} b \equiv \bigvee_{t=1}^{|S|-1} a \subseteq S_t^{t-z'} \wedge b \subseteq S_{t+1} \wedge b \not\subseteq S_t^{t-z'} \quad (6)$$

$$a \xrightarrow{0} b \equiv a \rightarrow b \quad (7)$$

The purpose behind this generalization is to account for large gaps of time between two events. For instance, if an event occurs in S_1 and is not recorded thereafter, can it be said to be associated with an event a time τ ? With windowed initiation $a \xrightarrow{\tau} b$, this question can be answered.

It is trivial to show that the set of initiation z -windowed rules is a subset of the set of all initiation rules: the set of rules $a \xrightarrow{0} b$ are equivalent to $a \rightarrow b$ and is vacuously a subset, and because $S_t^{t-z'} \subseteq S_t^*$ by definition, all initiation rules for $z > 0$ are also initiation rules. Thus, the rule $a \xrightarrow{z} b$ implies $a \rightarrow b$.

Initiation rule mining

We elicit initiation rules de novo using a method we call *initiation rule mining* (IRM). IRM is similar in design to ARM: candidate rules are proposed, then based off of their support in a transaction database, are assessed for their validity. The primary difference is that rather than looking within the same transaction, IRM mines rules by looking between different transactions contained in a single history.

Let T represent a set of histories (the transaction database) and $S \in T$. One possible incarnation of T (the incarnation used by GatewayNet) is illustrated in Table 1. Each record within the table is a triple (i, t, S_t) , such that $T_{i(t)} = S_t$.

Criteria must exist for candidate rules to be accepted or rejected, and several are traditionally used in ARM that apply here. Count Eq. 8 and support (Eqs. 9 and 10) are perhaps the most basic and may be used to filter out rules which run the risk of being statistically invalid [20, 24]; however, high limits may preclude many relationships from being discovered. Confidence is a measure of how likely the rule occurs when its antecedent occurs Eq. 11 and may be a more suitable measure for this purpose. Lift Eq. 12 is a measure of interest which considers the case where a and b are independent [20]. Finally, conviction is the frequency that the rule makes an incorrect prediction Eq. 13 [20]. Thresholds for inclusion are expressed as l_{count} , l_{sup} , l_{conf} , l_{lift} , l_{conv} , and h_{conv} respectively.

$$\text{count}(X) \equiv \sum_{S \in T} [X \subseteq S^*] \tag{8}$$

$$\text{sup}(X) = \frac{\text{count}(X)}{|T|} \tag{9}$$

$$\text{sup}(a \rightarrow b) = \text{sup}(a \cup b) \tag{10}$$

$$\text{conf}(a \rightarrow b) = \frac{\text{sup}(a \rightarrow b)}{\text{sup}(a)} \tag{11}$$

$$\text{lift}(a \rightarrow b) = \frac{\text{sup}(a \rightarrow b)}{\text{sup}(a) \times \text{sup}(b)} \tag{12}$$

$$\text{conv}(a \rightarrow b) = \frac{1 - \text{sup}(b)}{1 - \text{conf}(a \rightarrow b)} \tag{13}$$

The subset of candidate initiation rules for which these criteria met are called the set of *mined rules*. A rule is an element of the mined rules if and only if:

- $\text{count}(a \cup b) \geq l_{count}$
- $\text{sup}(a \rightarrow b) \geq l_{sup}$
- $\text{conf}(a \rightarrow b) \geq l_{conf}$
- $\text{lift}(a \rightarrow b) \geq l_{lift}$
- $l_{conv} \leq \text{conv}(a \rightarrow b) \leq h_{conv}$

As with ARM, the a priori principle may be used with IRM to reduce the number of candidates that must be considered when testing a proposed initiation rule for inclusion. An item set X is considered frequent if a) $\text{count}(X) \geq l_{count}$ and b) $\text{sup}(X) \geq l_{sup}$. Let $d_{max} \in \mathbb{Z}^+$ be the maximum degree for which to mine rules. Thus, rule proposal can be implemented as shown in Fig. 1, where $X \otimes Y$ is the outer product of X and Y .

Gateway rules

Recall that the gateway hypothesis predicts that the probability that b will arise out of a is greater than the probability that it would happen due to some other circumstance. When we say this, we say that a is a gateway into b and denote that relationship using $a \rightsquigarrow b$.

An initiation rule is known as a gateway rule (denoted $a \rightsquigarrow b$) whenever the probability that $a \rightarrow b$ Eq. 14 is greater than the probability that any combination of the remaining antecedents will initiate b . This is equivalent to positing that a (either directly or indirectly) causes b .

$$p(a \rightarrow b) = \frac{\text{sup}(a \rightarrow b)}{\text{sup}(b)} \tag{14}$$

A simple way of ensuring this condition is to calculate the proposed rule's *certainty* Eq. 1. The condition $\text{cert}(a \rightarrow b) = 1$ means that the probability that the subsequent arose out of a is precisely 50%, or alternatively that 50% of the remaining instances arose out of $a \nrightarrow b$. Therefore, by Eq. 1, we posit that a is the most likely cause of b when $\text{cert}(a \rightarrow b) > 1$. When the limit of $p(a \rightarrow b)$ approaches 1, $\text{cert } a \rightarrow b$ approaches ∞ : absolute certainty means that we posit b arises only from a Fig. 2

$$\text{cert}(a \rightarrow b) = \frac{p(a \rightarrow b)}{1 - p(a \rightarrow b)} = \frac{\text{sup}(a \rightarrow b)}{\text{sup}(b) - \text{sup}(a \rightarrow b)} \tag{15}$$

This test is necessary (albeit not sufficient) for the assertion that a given event is the singular cause of another. Even in [1], this degree of causation is not predicted: cigarettes or alcohol leads to cannabis. This method could only be used to therefore test the idea that cannabis singularly leads to other illicit drugs.

To test the most general form of the gateway hypothesis, it must be the case that the association occurs by greater probability than chance alone. Thus, we suggest

Table 1 A sample transaction database

ID	Time	Itemset	ID	Time	Itemset
1	0	{1, 2}	4	0	{1}
1	1	{3}	4	2	{2, 3}
1	4	{2, 3}	4	3	{1, 2}
2	0	{2}	5	0	{1}
3	2	{1}	5	1	{1, 2}
3	5	{2}	5	4	{1, 2, 3}

```

propose(T, z):

F0 = frequent item sets in E
R = ∅
foreach i from 1 to dmax:
    Fi = frequent item sets in (Fi-1 ⊗ Fi-1)
    foreach a in Fi:
        foreach b in Fi:
            next if a ⊆ b
            R ← R ∪ {a  $\xrightarrow{z}$  b}
return R
    
```

Fig. 1 Rule proposal algorithm using the a priori principle

that that gateway rules can be established using the condition $\text{cert}(a \rightarrow b) > l_{\text{cert}}$, which is the maximum certainty for which we will reject $a \rightarrow b$ as causal. To satisfy this hypothesis, the l_{cert} must be at least the threshold where we would admit chance occurrence Eq. 16.

$$l_{\text{cert}(\ast)} = \frac{1/|E|}{1 - 1/|E|} = \frac{1}{|E| - 1} \tag{16}$$

Consider the transaction database in Table 1; if one calculates the count for all of the item sets and degree one rules in the transaction database, then the values provided in Table 2 can be used to calculate support; for instance, I_1

has a support of 0.8 because it is involved in 4 or 5 histories. Initiation rule support can be calculated by finding all histories where Eq. 4 holds; because of this, $I_1 \rightarrow I_3$ has a support of 0.6 (Table 2).

Using this table, it is possible to derive the aforementioned metrics: for instance, $\text{lift}(I_1 \rightarrow I_3) = (0.6)/(0.8 \times 0.6) = 1.25$. To determine whether or not this $I_1 \rightarrow I_3$ is also a gateway rule, one calculates $\text{cert}(I_1 \rightarrow I_3) = (0.6)/(0.6 - 0.6) = 0.6/0$. Although this value is undefined, it can be interpreted as approaching ∞ ; thus, $I_1 \rightsquigarrow I_3$ can be said to hold.

Likewise, $\text{cert}(I_1 \rightarrow I_2) = (0.6)/(1 - 0.6) = 1.5$, so $I_1 \rightsquigarrow I_2$ in an unwindowed context. Let us now form initiation

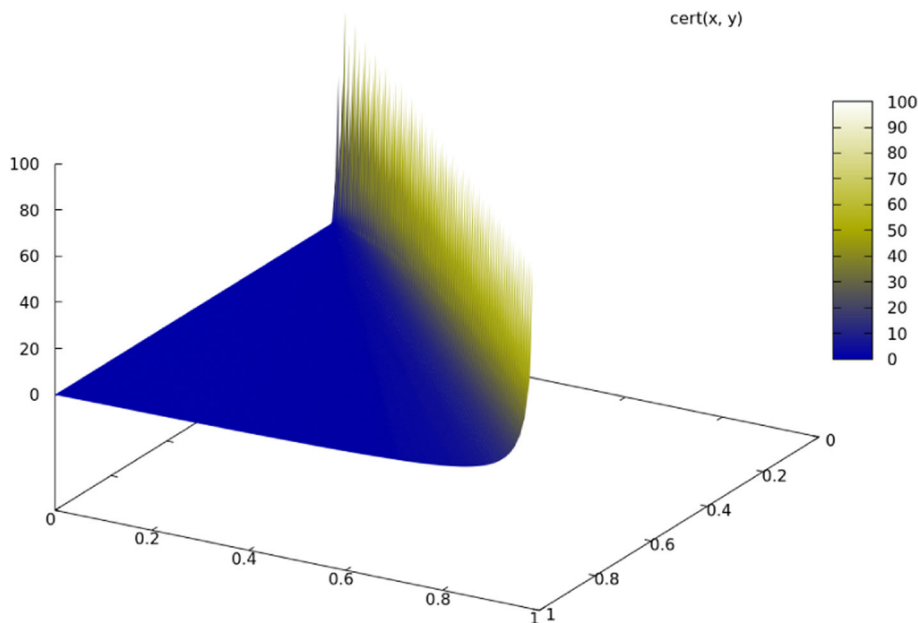


Fig. 2 The domain of $\text{cert}(x \rightarrow y)$. Certainty approaches ∞ (i.e., becomes absolute) as $\text{sup}(x \rightarrow y)$ approaches $\text{sup}(y)$. Color gradient represents certainty value for $\text{cert}(x \rightarrow y)$

Table 2 Counts and supports for all item sets and unwindowed initiation rules of degree 1 in Table 1

I.S.	Count	Sup.	I.S.	Count	Sup.
l_1	4	0.8	$l_1 \rightarrow l_2$	3	0.6
l_2	5	1	$l_1 \rightarrow l_3$	3	0.6
l_3	3	0.6	$l_2 \rightarrow l_3$	2	0.4
$\{l_1, l_2\}$	4	0.8	$\{l_2, l_3\}$	3	0.6
$\{l_1, l_3\}$	3	0.6	$\{l_1, l_2, l_3\}$	3	0.6

rules over window $z = 2$ (Table 3). Because the history for ID 3 does not initiate l_2 within the window, the support for $l_1 \xrightarrow{2} l_2$ drops to 0.4 and $\text{cert}(l_1 \xrightarrow{2} l_2) = (0.4)/(1-0.4) = 0.\bar{6}$. Thus, $l_1 \rightsquigarrow l_2$ because the certainty of $\text{cert}(l_1 \xrightarrow{2} l_2) \leq 1$ and therefore does not meet Eq. 1.

Visualization

The final phase that GatewayNet performs is visualization. This produces a directed graph which depicts relationships between initiation rules. Let G be a weighted digraph $G = \langle E', R \rangle$, where events E' constitute the graph's vertices, and rules R constitute edges between events. Let $r \in R$ be a quadruple such that $r \equiv \langle e_1 \in E', e_2 \in E', w, c \rangle$. Then for rule $a \rightarrow b$, vertices $\{a, b\} \in E'$, edge $r_{a \rightarrow b}$ is defined by Eq. 17, and membership of $a \rightarrow b$ in G is defined by Eq. 18.

$$r_{a \rightarrow b} = \langle a, b, \text{sup}(a \rightarrow b), \text{cert}(a \rightarrow b) \rangle \tag{17}$$

$$(a \rightarrow b) \in G \equiv \langle a, b, *, * \rangle \tag{18}$$

Synthetic data

To better characterize GatewayNet's behavior, we created a synthetic data set (Additional file 2) for which interaction is well characterized. This data set is explicitly constructed so that a complete history is available for each subject in the data set.

The synthetic data was generated according to a mathematical model described in the following paragraphs. This was done for the purposes of validation; although we also tested against empirical data, it is important that we verify

Table 3 Counts and supports for all item sets and initiation rules for $z = 2$ and of degree 1 in Table 1

I.S.	Count	Sup.	I.S.	Count	Sup.
l_1	4	0.8	$l_1 \xrightarrow{2} l_2$	2	0.4
l_2	5	1	$l_1 \xrightarrow{2} l_3$	3	0.6
l_3	3	0.6	$l_2 \xrightarrow{2} l_3$	2	0.4
$\{l_1, l_2\}$	3	0.6	$\{l_2, l_3\}$	3	0.6
$\{l_1, l_3\}$	3	0.6	$\{l_1, l_2, l_3\}$	3	0.6

that GatewayNet is well-behaved. The forthcoming model describes a population which is fixated on events it considers preferential, but allows for experimentation with other events.

Let E consist of events $\{I_1, I_2, \dots, I_n\}$, where $n \in \mathbb{Z}^+$, $E_0 \equiv \{\epsilon\} \cup E$, $e_1 \in E_0$, and $e_2 \in E_0$. A Markov chain of order 1 P (Additional file 1) is randomly constructed to represent transition probabilities from e_1 to e_2 (Table 4). Two real parameters are provided: the affinity f and interest s . Affinity represents the probability that a subject will be satisfied with e_1 and will ensure that the event occurs at time $t + 1$. Interest is a weight that represents the likelihood that the subject would independently ensure e_2 will occur. A special event, ϵ , represents the null event, which represents a transition from no event.

Naturally, each row in P must add to exactly 1.0; however, care must be taken to ensure that this criterion is met. Let P' represent an $|E_0| \times |E_0|$ matrix. Each element of P' is populated using Eq. 19: an event's self-transition $e_1 \rightarrow e_1$ is simply represented by its affinity, while any other transition is randomly distributed from the remaining probability. Because the row sum may not add up to exactly 1, each element is then normalized across the row Eq. 20.

$$P'(e_i \rightarrow e_j) = \begin{cases} a_i & i = j \\ \text{rand}[0, 1 - s_j P(e_i \rightarrow e_{j-1})] & i \neq j \end{cases} \tag{19}$$

$$P(e_i \rightarrow e_j) = \frac{P'(e_i \rightarrow e_j)}{\sum_{k=1}^{|E_0|} P'(e_i \rightarrow e_k)} \tag{20}$$

Because the generated history is considered to be a complete one, the initial state is always ϵ . Thus, time-point $t = 1$ is considered to be the first opportunity for which an initiation event can occur. At each time-point, between two and three initiations may occur. Each history may have up to 12 records in it; in total, we generated 56,578 simulated transactions over 8192 histories. Most of the events had support above 10% (Table 5). In total, 29,412 events were generated, corresponding to an average of 2.49 events per history.

We generated two gateway networks for the synthetic data: one for $l_{sup} = 0.20$ (Fig. 3), and one for $l_{sup} = 0.025$

Table 4 A sample Markov chain of order 1 describing transition probabilities between events $\{\epsilon, l_1, l_2, l_3\}$

	ϵ	l_1	l_2	l_3
ϵ	0.750	0.050	0.120	0.080
l_1	0.900	0.010	0.045	0.045
l_2	0.250	0.500	0.550	0.160
l_3	0.750	0.050	0.120	0.08

Table 5 Counts and supports for all 1-sets in the synthetic data

I.S.	Count	Sup.	I.S.	Count	Sup.
l_1	7230	0.9117	l_7	2891	0.3646
l_2	2260	0.2850	l_8	1508	0.1902
l_3	7526	0.9491	l_9	989	0.1247
l_4	1304	0.1644	l_{10}	1113	0.1404
l_5	5632	0.7078	l_{11}	661	0.0834
l_6	3790	0.4763	l_{12}	220	0.0277

(Fig. 4). In both instances, $l_{conf} = 0.5$, and $l_{lift} = 1$. Versions of the network without gateway rule highlighting, with gateway rule highlighting, and just the gateway rules were generated. Additionally, gateway networks were generated with window sizes of $z = 1$, $z = 2$, and $z = 3$ (Fig. 5). This was done to determine whether windowing had an effect on the synthetic data.

Empirical Data

Synthetic data is useful for evaluating the performance of GatewayNet since it is expected that some structures should arise within the output (therefore providing a method of validation). However, it should be noted that synthetic data does not necessarily model the real world; to test performance in that environment, an empirical data set was used.

UDS data obtained from 71,312 patient between August 1998 and June 2011 over nearly 111,359 emergency room visits at LSU Health Sciences Center (the hospital portion now belongs to University Health) in Shreveport, LA. This hospital is a Level I trauma center that serves

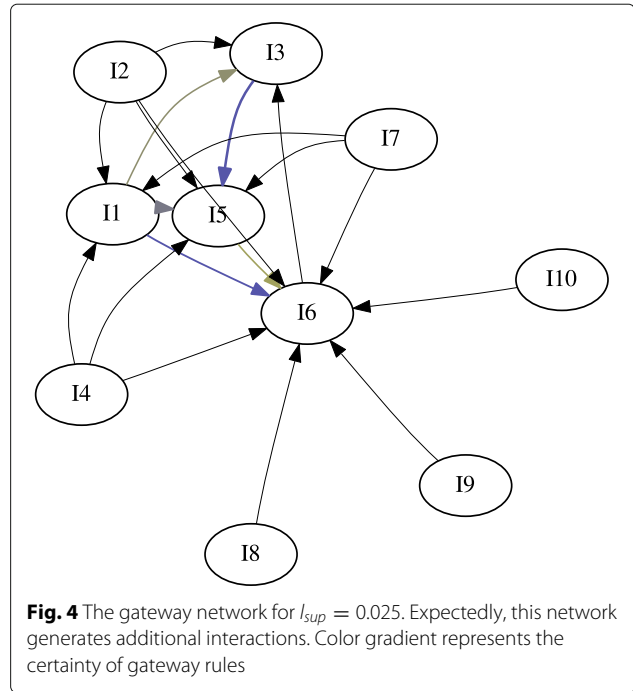


Fig. 4 The gateway network for $l_{sup} = 0.025$. Expectedly, this network generates additional interactions. Color gradient represents the certainty of gateway rules

the 7 parishes in LERN Region 7 (including the Shreveport/Bossier City area) [25]. Because Caddo Parish (where Shreveport resides) is adjacent to both the Texas and Arkansas borders, patients from east Texas and southern Arkansas are also frequently served.

During the screening interval, four screening panels were used, and during this time, some drugs

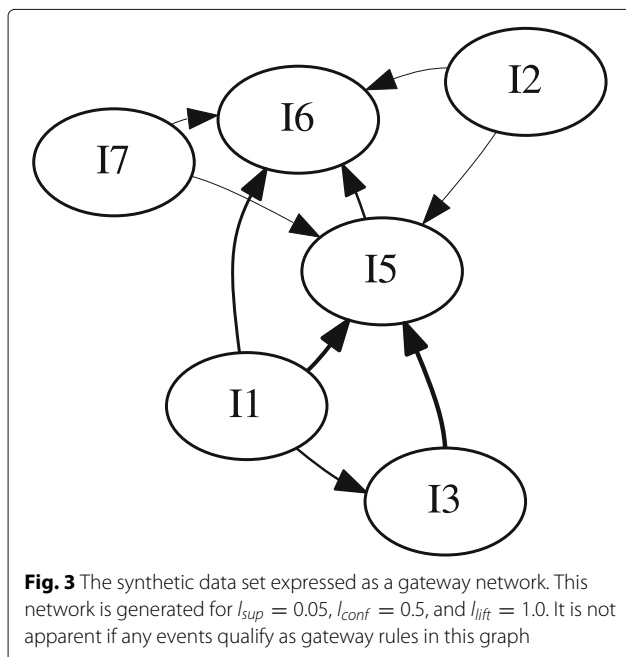


Fig. 3 The synthetic data set expressed as a gateway network. This network is generated for $l_{sup} = 0.05$, $l_{conf} = 0.5$, and $l_{lift} = 1.0$. It is not apparent if any events qualify as gateway rules in this graph

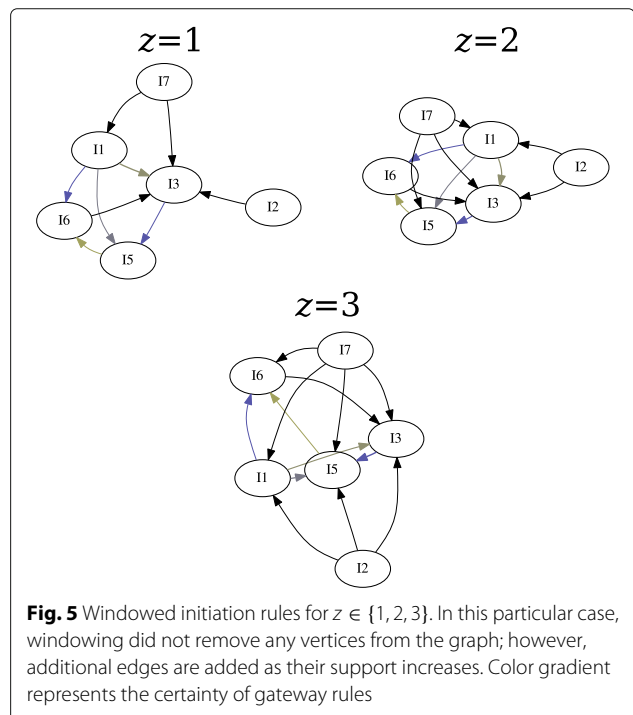


Fig. 5 Windowed initiation rules for $z \in \{1, 2, 3\}$. In this particular case, windowing did not remove any vertices from the graph; however, additional edges are added as their support increases. Color gradient represents the certainty of gateway rules

were not tracked consistently. These drugs were: 3,4-methylenedioxymethamphetamine (MDMA or ecstasy) and methadone (tested during 2007–2011), methamphetamine and propoxyphene (1998–2000, 2002–2004), and barbiturates (1998–2007).

Prior to processing, we removed demographic data and then assigned each patient a random identifier (ID) by first shuffling the list of patients, then assigning each patient in the shuffled list a sequential ID. Additionally, screening dates were converted to their corresponding Lilian day number. The day number was then scaled by 1440 (the number of minutes in the day) and the time of screening in minutes was added to the date. Finally, each patient’s screening time was calibrated to the first by subtracting the first screening’s timestamp.

This was done for to ensure that the screening time is expressed as an integer. Additionally, because methamphetamine and MDMA are amphetamines and methadone is an opiate, any instance of either was converted to this category prior to any processing. Because many patients only visited once or did not test positive for any drugs, we restricted the list of histories to those with at least two time-points and at least one positive result. Finally, a history was only accepted if there was at least one more time-point following the time-point of the first positive result. In total, 11,364 histories over 42,745 time-points remained.

This data was first processed using unwinded IRM (Fig. 6). We set the parameters $l_{count} = 30$, $l_{sup} = 0$, $l_{conf} = 0.25$. Minimum count was used instead of support because of the relatively few number of histories involving drug use (Table 6).

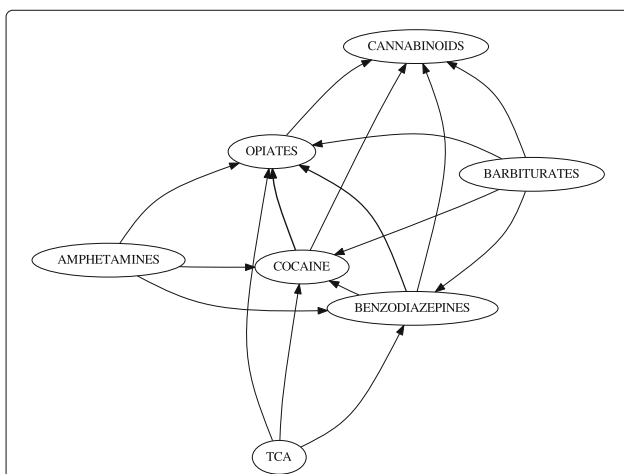


Fig. 6 Gateway Network for LSUHSC-S data. No gateway rules were found; however, the initiations TCA → BENZODIAZEPINES and BARBITURATES → BENZODIAZEPINES were discovered. Additionally, benzodiazepines, barbiturates, cocaine are seen to initiate cannabinoids

Table 6 Counts and supports for all 1-sets in the LSUHSC-S data

I.S.	Count	Sup.
Amphetamines	1459	0.1284
Barbiturates	666	0.0586
Benzodiazepines	4120	0.3625
Cannabinoids	5937	0.5224
Cocaine	3822	0.3363
Methadone	407	0.0358
Opiates	4525	0.3982
Phencyclidine	114	0.0100
TCA's	350	0.0308

In addition to performing unwinded mining, we mined initiation rules within a window of 525,600 minutes (1 year) (Fig. 7). This was done to remove rules which were primarily supported by spurious positives. Opiates were sometimes administered to incoming patients or as a result of emergency surgery. Because of usage this arising from medical intervention rather than choice, we further removed rules of the form $x \rightarrow$ OPIATES (Fig. 8).

Results

Because we evaluated two data sets, we discuss the results for each data set separately. In the following subsection, we will discuss the results of applying GatewayNet to the synthetic data set. Afterwards, we discuss the results with respect to the empirical data from LSUHSC-S.

Synthetic data

A relatively simple network with multiple interactions was generated (Fig. 3). It would appear that with our synthetic data set, event I_1 initiates I_3 , I_5 and I_6 , while I_5 initiates I_6 and I_3 initiates I_5 . In this data set, I_1 , I_3 , and I_5 are gateway events.

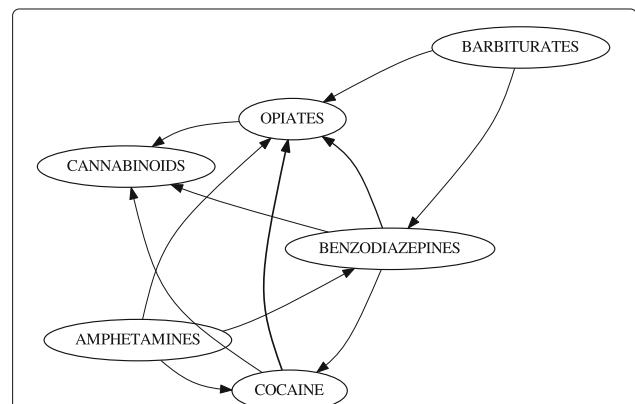
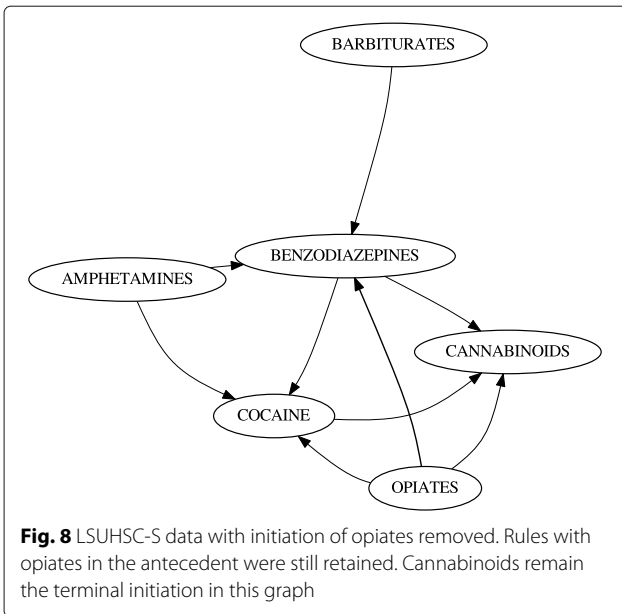
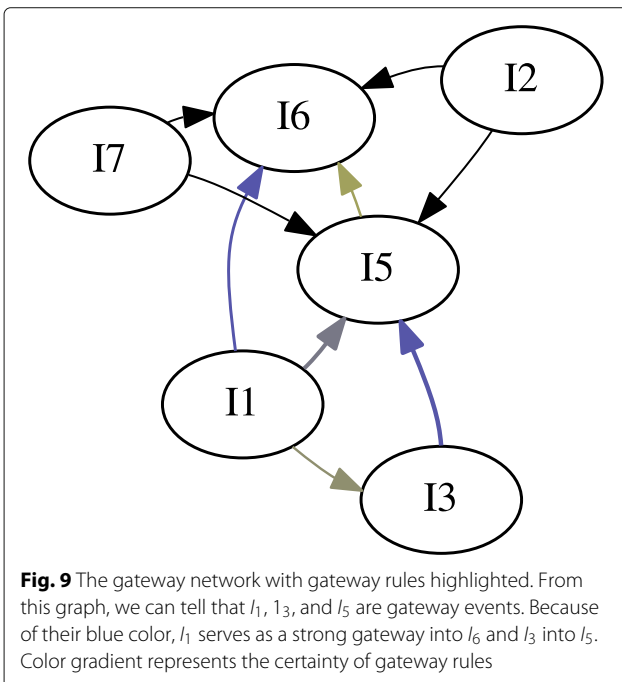


Fig. 7 LSUHSC-S data over a year-long window. TCAs were eliminated when windowing was applied, and barbiturates no longer directly initiate cannabinoids



Inspecting Fig. 3, there appear to be no clear gateway; however, this is misleading. The set of actual gateway events predicted by our algorithm are I_1, I_3 , and I_5 , (Fig. 9). This may not be very surprising: these are also the most frequent events (Table 5). This can be more easily seen by removing the nodes which do not correspond to gateway rules (Fig. 10).

It can be clearly seen that $I_3 \rightsquigarrow I_5$ with a high degree of certainty (32.3254). However, it is also true that $I_3 \rightsquigarrow I_5$



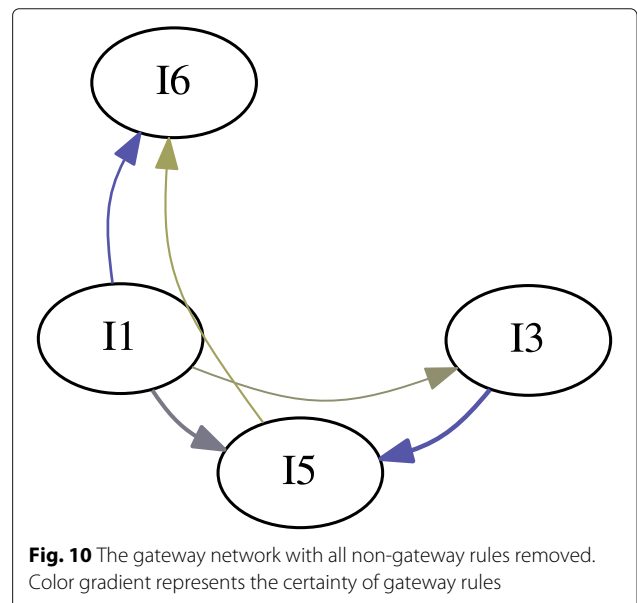
(albeit with a weaker certainty of 19.4058) Fig. 5. Additionally, Both I_1 (32.8393) and I_5 (4.0736) are gateways into I_6 . The 2nd degree rules $\{I_1, I_3\} \rightarrow I_5$ and $\{I_1, I_5\} \rightarrow I_6$ were mined.

In this case, windowing did not effect the vertices in the gateway network; however, the edges reported did change (Fig. 5). Notably, the number of edges associated with I_2 and I_7 increased with the window size. However, some of the rules associated with these events are not mined in the unwindowed sample, possibly due to diminishing support as maximum count increases.

Empirical data

No gateway rules were observed in the data; however, some trends could be observed. In particular, tricyclic antidepressants (TCAs) and barbiturates both initiate benzodiazepine use (Fig. 6). This is not surprising; TCAs and barbiturates were once regularly prescribed, but have been replaced benzodiazepines (which have fewer risks). Although this does not qualify as a gateway event, it confirms that a known initiation even can be captured; the TCA relationship is also filtered out in the year-long window (Fig. 7).

Curiously, cocaine, opiates, and benzodiazepines initiate cannabinoids according to the data (Figs. 6, 7). In addition to barbiturates, amphetamines also initiate benzodiazepines (Figs. 6, 7). Before filtering out initiation rules involving opiates in the subsequent, cocaine, benzodiazepines, and amphetamines were also found to initiate opiates (Figs. 6, 7 and 8). The only drugs removed between the unwindowed and windowed variants were TCAs (Fig. 7).



Discussion

As with the “Results” section, this section discusses the results for the synthetic data and empirical data separately. In the following subsection, we will discuss whether or not mining the synthetic data yielded the expected results. Afterwards, we discuss observations noted with respect to mining the LSUHSC-S UDS data.

Synthetic data

The synthetic data showed a peculiar phenomenon: because $l_{cert} = 1$, it might be expected that one (and only one) event may serve as a gateway into another. As it turns out, this expectation is unwarranted: a high certainty means that a may be necessary to explain an event, not that it is sufficient to do so. In this case, two hidden gateway rules $\{I_1, I_3\} \rightsquigarrow I_5$ and $\{I_1, I_5\} \rightsquigarrow I_6$ were discovered. But what does this mean?

Let us first consider $\{I_1, I_3\} \rightsquigarrow I_5$. One possibility is that I_1 and I_3 are co-requisite for the event. Consider the model from which the synthetic data is derived. In this model, an event for which there is high affinity will be quickly retained and will become recurrent; however, there is no reason that this might be the only event to occur within that time-point. Because of this, a history will show frequent experiments with other events: each event is an avenue for other events to occur alongside it.

With respect to drug use, this model is perhaps pessimistic: it predicts co-usage of one drug with other drugs, even when the user has high affinity with another one. In our model, frenetic experimentation seems to occur as the search for other events with high affinity continues. Is this a reasonable model of drug use? It is known that comorbidities in drug use often exist in reality. In Australasian countries, rates of experimentation of around 40% have been observed [11, 26].

If this does reflect trends in drug use, then it is an interesting result, as our method would be able to detect this phenomenon. In the synthetic data, $I_1 \rightarrow I_5$ at a probability of 95.099%, whereas $I_3 \rightarrow I_5$ with a probability of 96.999%. This concedes the possibility that they are used in combination.

But as it turns out, that is not the only possible explanation, because without any windowing, a gateway rule may be formed so long as I_1 and I_2 precede I_5 at *any time in the past*. We should also note that $I_1 \rightsquigarrow I_2$ with a probability of 94.069%. It is possible that I_3 is directly responsible for the transition into I_5 , and in reality, it is probably so: the first-order Markov chain used to model the relationships between events cannot explicitly express $\{I_1, I_3\} \rightarrow I_5$, nor does it actually encode $I_1 \rightarrow I_3 \rightarrow I_5$ because it lacks the required history. Because of this, the latter is probably a better explanation.

Empirical data

The LSUHSC-S data did not exhibit any gateway rules, and there are several potential reasons for this. It could simply be that the gateway hypothesis does not manifest itself in the population (or at least this sample). This represents acceptance of the null hypothesis that the drugs involved do not progress in development stages as predicted by Kandel and Yamaguchi [1–3].

However, it must be stated that there is another source of error which is likely to be present in the data: the drug screenings are collected during trauma center visits, and this offers an incomplete usage history of each patient. In many countries, illicit drug use is a criminal offense, and even if it were not, drug use is commonly voluntary and this means that intoxication is probably a desirable state for the user. Because of this, we can expect that patients will not seek the help of the trauma center merely because they have consumed an illicit drug; instead, we would expect to see that these patients will do so because of the perceived risk of dying or due to circumstantially related incidents.

Because of this, we do not know the patient's drug history between intervention, and there may be biases due to the time it takes to clear metabolites from their systems. It is known, for instance, that some drug metabolites are excreted at different rates than others [27]. The cannabinoid metabolite 11-nor-9-carboxy- Δ^9 -tetrahydrocannabinol (THC-COOH) has a urinary half-life of about two days [28] compared to the 7.5 hour half-life of the cocaine metabolite benzoylecgonine [29]. This may mean that some initiation events will not be captured, particularly whenever screenings are separated by months or years.

Additionally, it was previously mentioned that all of the drugs tested by the panel (with the exception of cannabinoids) had a recognized medicinal use at the time. This true of opiates (for instance) to such a degree that initiations into opiates had to be filtered out of our data. However, we do not know if or when certain drugs were administered legitimately to patients (e.g., via prescription or surgical intervention prior to screening), and had gateway rules been found, they would have been suspect because of this.

We did, however, observe what is very likely to be this phenomenon in action. By finding that barbiturates and TCAs initiate benzodiazepines (Fig. 6), we were able to observe a known transition in medical practice. This initiation event was lost when initiation rules were limited to a year-long window (Fig. 7); since barbiturates and TCAs and benzodiazepines are antidepressants and anxiolytic respectively, it may have been the case that these were administered to patients en route as the result of psychiatric intervention. One may therefore predict that additional emergency intervention was simply not

required within this window.

By removing rules with opiates in the rule's subsequent, we notice that the transition into cannabinoids remains (Fig. 8). Interestingly, this is suggested by Kandel because she initially hypothesizes that the association is bidirectional [1]; in fact, we generally see this prediction also holding for benzodiazepines and cocaine. However, it must be restated that we did not find any gateway rules and that this is association. One potential explanation for this is that cannabinoids may have been easier to obtain illicitly over the sampling period than alternatives.

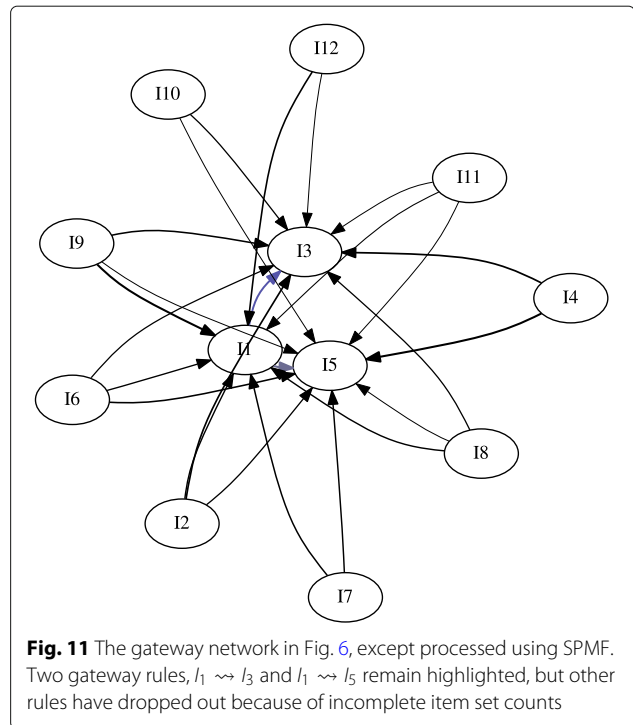
Comparison to existing software

Several implementations of SRM are mentioned in the "Discussion" section which may be compared to GatewayNet. As far as we are aware, no software utilizes the certainty measure, so any calculation thereof is extrinsic to other SRM software. However, a comparison can be made with existing software as long as support information of each item set mined is available.

To supply the other algorithms, we utilized SPMF, which implements both the SPADE and SPAM algorithms [30]. SPMF takes slightly different input than GatewayNet, and the input data was converted by assigning item labels to integer values and by combining each history into a single line. As SPMF does not have an option to set minimum count, this was achieved by using setting minimum support to 0.0005166 (30/58067 transactions). Both SPAM and SPADE produced equivalent output that varies only in output order and consists of 56,767,617 individual item sets. Because of this, the remaining analysis occurs on the SPAM output.

Unsurprisingly, both algorithms calculate the same number of 1-sets as GatewayNet does; however, special care must be taken to make the results comparable to GatewayNet's because SPMF outputs its item sets as a series of time-points. In order to perform this comparison, we first "trimmed" SPMF's item sets so that the repeated sequences at either extrema are truncated; then, each such item set with the maximum support is counted. In total, this reduced the number of item sets to 16,415,526.

From this set, we calculated all frequent subsets in the data. This was first done by mining all subsets of length k for each SPMF item set S , for $0 < k < |S|$. A ruleset was then generated with a minimum support of 0.025% and propagated to GatewayNet's visualization software, gatewaynet-links. We found that no rule had a lift of 1 or greater; the graph has also reduced the set of gateway rules to $I_1 \rightsquigarrow I_3$ and $I_1 \rightsquigarrow I_5$ (Fig. 11). This might be explained by a loss of records due to SPMF's output. In that output, $\{A, B\}$ represents a single *transaction* where this set of events occurs, and it is distinct from $\{\{A\}, \dots, \{B\}, \dots\}$ or $\{\{B\}, \dots, \{A\}, \dots\}$. In contrast, GatewayNet considers such histories to correspond to the item



set $\{A, B\}$. Because of this, many small sequences of low support may be culled by the support threshold, therefore altering each item set's frequency. Nonetheless, using SPMF as a basis for item set mining approaches the results of GatewayNet.

Conclusions

IRM (as implemented by GatewayNet) shows promise to demystify the Gateway Hypothesis, but it may also be useful in the prediction of any event (as our synthetic set demonstrates). The quality of data provided to GatewayNet will strongly affect its output; however, with good data, it may not only be able to highlight initiation events, but also actual gateway events as well.

The ability of GatewayNet to predict initiation rules is expectedly dependent on data quality. In our case, emergency room UDS yielded no support for the gateway hypothesis; however, due to the nature of the screening, a full patient history might not be available.

It should be noted that GatewayNet is not presently optimized for speed. The a priori algorithm is known to be sub-optimal for association rule mining, and other SRM algorithms have explored optimization techniques regarding speed. In the future, we would like to explore techniques such as FP-growth as avenues for improving runtime speed.

Availability and requirements

Project name: GatewayNet

Project home page: <https://sun.cs.lsus.edu/software/>

gatewaynet/

Operating systems: GNU/Linux (amd64/x86_64)

Programming language: C++98/Perl

Other Requirements: Perl 5.22.1, File::Basename, Getopt::Long, GraphViz

License: Non-commercial Use

Any restrictions to use by non-academics: commercial use must be licensed

Additional files

Additional file 1: Supplementary Data. This file contains the transaction database for our synthetic data. (TXT 702 kb)

Additional file 2: Supplementary Data. This file contains the Markov chain used to generate `sample.txns.txt` (CSV 3 kb)

Abbreviations

ARM: Association rule mining; BBN: Bayesian Belief Network; ED: Emergency department; ID: Identifier; IRM: Initiation Rule Mining; LSUHSC-S: LSU Health Sciences Center in Shreveport; MDMA: Methylenedioxymethamphetamine; PCP: phencyclidine; SRM: Sequential Rule Mining; TCA: tricyclic antidepressant; THC: Δ^9 -tetrahydrocannabinol; THC-COOH: 11-nor-9-carboxy- Δ^9 -tetrahydrocannabinol; UDS: urine drug screening

Acknowledgments

We would like to thank M. Ethan Black for his assistance in ensuring the comprehension of this document.

Funding

Research reported in this manuscript was supported by the Noel Foundation and Louisiana Board of Regents Endowed Professor/Chair Programs: Abe Sadoff Chair in Bioinformatics and Lisa Burke Bioinformatics Scholarship. This research was supported in part by grants from the National Institute of General Medical Sciences of the NIH under Award Numbers P30GM110703 and P20GM103424.

The funding bodies had no role in the design of the study and collection, the analysis, interpretation of the data and in writing the manuscript.

Availability of data and materials

Both the transaction database and Markov chain for the synthetic data used in this article are included in this published article. Due to ethical constraints related to patient privacy, the transaction database for the empirical data used in this article is not available.

Authors' contributions

PCSRK, MT, and UC implemented GatewayNet. NK is a domain expert and provided consultation and empirical data required for development. TA collected the empirical data and provided valuable advice in results interpretation. PCSRK, MT, NK, and UC drafted the manuscript. All authors wrote and approved the final manuscript. All authors have read and approved this manuscript.

Ethics approval and consent to participate

Approval to use patient data has been provided by the LSU Health Institutional Review Board. All investigators have received mandatory certification from the Collaborative Institutional Training Initiative and understand the ethical responsibility in working with human subjects' data. The protocol for human studies was approved by the Louisiana State University Health Science Center, Institutional Review Board, an IRB protocol number H12-151.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Computer Science, LSU Shreveport, 1 University Place, 71115 Shreveport, USA. ²Department of Pharmacology, Toxicology and Neuroscience, LSU Health Shreveport, 1501 Kings Highway, 71103 Shreveport, USA.

³Department of Emergency Medicine, LSU Health Shreveport, 1501 Kings Highway, 71103 Shreveport, USA. ⁴Center for Molecular and Tumor Virology, LSU Health Shreveport, 1501 Kings Highway, 71103 Shreveport, USA.

Received: 10 July 2018 Accepted: 4 April 2019

Published online: 23 April 2019

References

1. Kandel D. Stages in adolescent involvement in drug use. *Science*. 1975;190(4217):912–4.
2. Yamaguchi K, Kandel D. Patterns of drug use from adolescence to young adulthood: li sequences of progression. *Am J Pub Health*. 1984;74(7):668–72.
3. Kandel D, Yamaguchi K. From beer to crack: Developmental patterns of drug involvement. *Am J Pub Health*. 1993;83(6):851–5.
4. O'Donnell JA, Clayton RR. The stepping-stone hypothesis – marijuana, heroin, and causality. *Chem Depend*. 1982;4(3):229–41.
5. Baumrind D. Familial antecedents of adolescent drug use: A developmental perspective. In: Jones CL, Battjes RJ, editors. *Etiology of Drug Abuse: Implications for Prevention*. 1st ed. Rockville: Department of Health and Human Services; 1985. p. 13–44.
6. Baumrind D. Specious causal attributions in the social sciences: the reformulated stepping-stone theory of heroin use as an exemplar. *J Personality Socl Psych*. 1983;45(6):1289–98.
7. Vanukov MM, et al. Common liability to addiction and "gateway hypothesis": Theoretical, empirical, and evolutionary perspective. *Drug Alcohol Depend*. 2012;123(1):3–17.
8. Panlilio LV, Zanetti C, Barnes C, Solinas M, Goldberg SR. Prior exposure to thc increases the addictive effects of nicotine in rats. *Neuropsychopharmacology*. 2013;38(7):1198–208.
9. Ellgren M, Spano SM, Hurd YL. Adolescent cannabis exposure alters opiate intake and opioid limbic neuronal populations in adult rats. *Neuropsychopharmacology*. 2007;32(3):607–15.
10. Cadoni C, Pisanu A, Solinas M, Acquas E, Chiara G. Behavioural sensitization after repeated exposure to Δ^9 -tetrahydrocannabinol and cross-sensitization with morphine. *Psychopharmacology*. 2001;158(3):259–66.
11. Fergusson DM, Boden JM, Horwood LJ. Cannabis use and other illicit drug use: testing the cannabis gateway hypothesis. *Addiction*. 2006;101(4):556–69.
12. Latkin CA, et al. Socially desirability response bias and other factors that may influence self-reports of substance use and hiv risk behaviors: A qualitative study of drug users in vietnam. *AIDS Educ Prev*. 2016;28(5):417–25.
13. Johnson TP. Sources of error in substance use prevalence surveys. *Intl Schol Res Notices*. 2014;2014(2014):.
14. Baden LR, Horowitz G, Jacoby H, et al. Quinolones and false-positive urine screening for opiates by immunoassay technology. *J Am Med Assoc*. 2001;286(24):3115–9.
15. Brahm NC, Yeager LL, Fox MD, Farmer K, Palmer TA. Commonly prescribed medications and potential false-positive urine drug screens. *Am J Health-Syst Pharm*. 2010;67(16):1344–50.
16. Eells E, Sober E. Probabilistic causality and the question of transitivity. *Phil Sci*. 1983;50(1):35–57.
17. Beenstock M, Rahav G. Testing Gateway Theory: do cigarette prices affect illicit drug use? *J Health Econ*. 2002;21(2):679–698.
18. Maldonado-Molina MM, Lanza ST. A framework to examine gateway relations in drug use: An application of latent transition analysis. *J Drug Issues*. 2010;40(4):901–24.
19. Bretteville-Jensense AL, Jacobi L. Climbing the drug staircase: A bayesian analysis of the initiation of hard drug use. *Psychol Med*. 2006;37(1):49–60.
20. Han J, Kamber M, Pei J. *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann; 2012.

21. Fournier-Viger P, Gomariz A, Campos M, Thomas R. Fast vertical mining of sequential patterns using co-occurrence information. In: Tseng VS, et al., editors. Proc. of the 18th Pacific-Asia Conf. on Knowledge Discovery and Data Mining. Switzerland: Springer International Publishing; 2014. p. 40–52.
22. Ayres J, Gehrke J, Yiu T, Flannick J. Sequential PAttern mining using a bitmap representation. In: Special Interest Group on Knowledge Discovery in Data. New York: Association for Computing Machinery; 2002.
23. Gomariz A, Campos M, Marin R, Goethals B. ClaSP: An efficient algorithm for mining frequent closed sequences. In: Pei J, et al., editors. Proc. of the 17th Pacific-Asia Conf. on Knowledge Discovery and Data Mining. Switzerland: Springer International Publishing; 2013. p. 50–61.
24. Brin S, Motwani R, Ullman JD, Tsur S. Dynamic itemset counting and implication rules for market basket data. In: Proc. of the 1997 ACM SigMOD int'l conf. on Management of Data. New York: Association of Computer Machinery; 1997. p. 255–64.
25. Louisiana Emergency Response Network. State Designated Trauma Centers. <http://lern.la.gov/trauma/state-designated-trauma-centers/>. Accessed 28 Feb 2018.
26. Agrawal A, Lynskey MT, Bucholz KK, Martin NG, Madden PAF. Contrasting models of genetic co-morbidity for cannabis and other illicit drugs in adult Australian twins. *Psychol Med.* 2006;37(1):49–60.
27. Verstraete AG. Detection times of drugs of abuse in blood, urine, and oral fluid. *Ther Drug Monit.* 2004;26(2):200–5.
28. Musshoff F, Madea B. Review of biologic matrices (urine, blood, hair) as indicators of recent or ongoing cannabis use. *Ther Drug Monit.* 2006;28(2):155–64.
29. Cone EJ, Weddington WW. Prolonged occurrence of cocaine in human saliva and urine after chronic use. *J Anal Tox.* 1989;13(2):65–68.
30. Fournier-Viger P, Lin CW, Gomariz A, Gueniche T, Soltani A, Deng Z, Lam HT. The SPMF open-source data mining library version 2. In: Tseng VS, et al., editors. Proc. of the 19th European Conference on Principles on Data Mining and Knowledge Discovery. Cham: Springer LNCS; 2016. p. 36–40.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

