Estimating The MBTI Of Twitter Accounts With Graph Neural Networks Over Neo4j

Georgios Drakopoulos[1,*], Eleanna Kafeza[2]

[1]*Department of Informatics, Ionian University, Tsirigoti Sq. 7, Kerkyra 49100, Hellas*
[2]*College of Interdisciplinary Studies, Zayed University, UAE*

Abstract

Intelligent agents are indispensable and flexible autonomous tools for efficiently mining large graphs for heterogeneous knowledge. Twitter is a prime case in point with structural and functional attributes such as original multimedia content posting and retweeting revealing important affective information about accounts. Additionally, this can be facilitated by including hashtag emotional polarity and reactions to political, social, or even historical events. Further insight can be gained by moving one step forward from individual emotional reactions to integrated personality estimations such as the MBTI taxonomy. An intelligent agent has been developed with a stochastic account visiting policy based on preferential attachment, an optional evolving forgetting factor for penalizing vertices appearing too frequently, and the capability to yield an MBTI estimate based on a graph neural network. The results indicate the superior performance of the proposed heuristic based on evaluation criteria including community size distribution and hashtag coherency.

1. Introduction

Intelligent agents (IAs) are autonomous digital entities with extensive capabilities for maintaining and ensuring the smooth operation of massive infrastructure, mostly networks of various types. Depending mainly on their technology and operating principles, an IA may have extended command and control capabilities while requiring only the initial programming in order to properly function, excluding of course any sort of necessary local input. Recently IA has advanced beyond a fixed set of rules to integrating various level of machine learning (ML) capabilities, provided sufficient computing power is available. The inclusion of neural networks architectures such as graph neural networks (GNNs) which take full advantage of the local network topology constitutes a major addition. Perhaps the most well-known representation of such an agent in pop culture, albeit possessing far more capabilities than those of its contemporary counterparts, is that of the iconic agent Smith from *The Matrix*[1].

Twitter mining analytics provide a significant opportunity to various organizations to improve both decision making processes and marketing performance. This is achieved as businesses are allowed through Twitter mining to gain invaluable insights on the dynamics and collective behavior of their customer base or any other online target group for that matter. In turn this yields more accurate predictions of key factors such as future demand, reactions to new products, or brand loyalty to name only a few. Twitter analytics tailored for this task include community structure discovery algorithms, hashtag flow analysis and information diffusion strategies, digital influence computation, and link prediction tools. Such insight is obtained frequently from the computationally challenging task of processing a diverse set of *follow* relationships, hashtags, or tweets. Such attributes are either of structural nature in the sense that they are about the social graph itself or functional as they pertain to the activity of the various entities, mostly the Twitter accounts, which use said graph.

Among the functional features the affective ones have recently garnered considerable research attention since emotions are the primary motivations behind human actions. To this end, attributes such as the emotional polarity of tweets and hashtags are considered as major indicators of how a Twitter account would react to various events and rely heavily on emotion models such as those proposed by Plutchik or Ekman. However, personality taxonomies such as the Myers-Brigs taxonomy indicator (MBTI) go beyond individual emotional responses and provide a more general framework for systematically evaluating sequences of account reactions as they take into consideration the higher cognitive functions driving them. For instance, personalities with an extrovert pre-

[1]https://www.imdb.com/title/tt0133093

disposition will be typically more vociferous compared to introvert ones for the same event.

Table 1
Notation Summary

Symbol	Meaning	First in
$\triangleq$	Equality by definition	Eq. (1)
$\{s_1, \ldots, s_n\}$	Set with elements $s_1, \ldots, s_n$	Eq. (2)
$(t_1, \ldots, t_n)$	Tuple with elements $t_1, \ldots, t_n$	Eq. (1)
$\langle s_k \rangle$	Sequence with elements s_k	Algo. 1
$|S|$	Set, sequence, or tuple cardinality	Eq. (4)
$(u, v; l)$	Edge from u to v with label l	Eq. (18)
$\Gamma_i(v)$	Inbound neighborhood of v	Eq. (7)
$\Gamma_o(v)$	Outbound neighborhood of v	Eq. (7)
$\text{prob}\{\Omega\}$	Probability of event Ω occurring	Eq. (7)
$S_1 \setminus S_2$	Asymmetric set difference	Eq. (8)
$\varphi(\cdot)$	Logistic function	Eq. (12)
$\psi(\cdot)$	Hyperbolic tangent function	Eq. (14)
$\langle g \| f \rangle$	KL divergence of g from f	Eq. (20)
$\mathbf{I}_n$	$n \times n$ identity matrix	Eq. (4)

The primary research objective of this conference paper is the development of an IA determining its next jump based on a strategy exploiting local structural information as well as the MBTI profile of the neighboring vertices. The latter is achieved by a preprocessing stage where a GNN performs personality type prediction based on the standard MBTI profiles. Each given vertex has a set of ground truth affective attributes which allow the deployment of multiple psychological interfaces depending on the application and the type of IA querying the vertex in question. Additionally, IAs take into consideration the MBTI personality of neighboring vertices prior to moving to one of them. The aforementioned factors differentiate significantly this conference paper from the vast majority of previous approaches.

The remainder of this conference paper is structured as follows. In section 2 the recent scientific literature regarding IAs, graph mining, and personality models are briefly reviewed. IA design is described in detail in section 3. The results obtained by the action of the proposed IA on benchmark graphs are examined in section 4, whereas possible future research directions are given in section 5. Sets are denoted by capital letters. Bold capital letters denote matrices, small bold vectors, and small scalars. Acronyms are explained the first time they are encountered in the text. In function definitions parameters are separated of the arguments by a semicolon. Finally, in table 1 the notation of this work is summarized.

2. Related Work

IA design relies on a number of approaches [1] as they have to function on their own in large digital infrastructure extending as a result the digital awareness of the organizations deployed them. IAs often have to take decisions [2, 3] which in turn rely on operational criteria based on aspects like anthropomorphism [4], maintaining trust with human users [5], cognitive functionality [6], connecting IAs with sensors [7], action explainability [8], and even the possible role of voice [9]. IAs can be employed in many capacities like protecting critical industrial cyber-physical infrastructure [10], communicating with humans through dynamic oral conversations [11], facilitating social interactions [12], recommending charging points for electric vehicles [13], modeling financial markets [14], and even shaping fashion trends [15]. In order for IAs to adapt to complex and nonstationary environments, ML capabilities have been recently added to them [16]. ML techniques cover a broad spectrum of options such as variational encoders [17], reinforcement learning [18], deep learning in various forms [19], and cooperative learning [20]. Possible extensions for use with IAs are tensor stack networks (TSNs) [21], GNNs [22], adversarial neural networks (ANNs) [23], and self organizing maps (SOMs) [24].

Graph mining extracts nontrivial knowledge from linked data [25]. For example, approximating directed graphs with undirected ones based on density criteria [26]. Community discovery for large graphs has taken many forms due not only to instance size [27] but also because many and equally valid graph community definitions exist [28, 29]. For instance, communities may well be build on trust [30], spatiotemporal patterns [31], social behavior [32], multiple connectivity criteria [33], noiseless patterns [34], spatial behavior akin to that of geolocation services [35], privacy preserving constraints [36], and simultaneous structural and affective criteria [37]. Applications include trajectory planning for autonomous race vehicles [38], political [39] and commercial [40] digital campaign designs, biomedical document recommendation based on a keyword-term-document tensor model [41], opinion mining [42], and assessing the affective resilience of Twitter graphs [43, 44].

Personality indicators [45] go beyond emotion models like the *emotion wheel* [46] or *big five* [47] since they provide a framework for predicting reactions to a wider array of stimuli [48], whereas the various emotion models only describe a single reaction. MBTI is used among others for designing digital campaigns on social media [49] perhaps in conjunction with other major attributes of the underlying domain [50]. Also MBTI taxonomy has been used to assess leadership traits [51]. Examples include exploring political Twitter [52], predicting brand loyalty [53], and analyzing the dynamics in social media communities [54]. Tensor distance metrics [55] can be used to cluster personalities in the MBTI space, especially when personalization is intended [56].

3. Intelligent Agent Design

3.1. Graph representation

Multilayer graphs in this work will be the algorithmic cornerstone for representing and analyzing the interaction between Twitter accounts and the IA as the latter will be circulating along the edges of such a graph. Multilayer graphs extend ordinary ones by including a set of distinct edge labels along with the requirement that an edge has exactly one and, thus, allowing as many as multiple edges between any two vertices as long as their labels are distinct. As such, concepts such as vertex degrees and paths have to be redefined to take into consideration this extension. As each label carries its own structural, functional, and semantic meaning, many underlying domains can be better modeled. Definition 1 formally introduces the class of labeled multilayer graphs.

Definition 1 (Labeled multilayer graphs). *A labeled multilayer graph G expresses simultaneous connections between its vertices allowing the formulation of higher order patterns and it is defined as the ordered triplet of* (1):

$$G \stackrel{\triangle}{=} (V, E, L) \tag{1}$$

The components of a multilayer G and the role they play in representing Twitter accounts and the associated interactions and activities are the following:

- The set of vertices V comprises of the entities the relationships are built on. In the context of the proposed methodology V consists of anonymized Twitter accounts.
- The set $E \subseteq V \times V \times L$ is the set of labeled edges. Therefore, not only does each edge have orientation but also a label. In this work they denote Twitter interaction.
- The label set L depends directly on the semantics of the underlying domain. In this particular case, L contains the four elements of equation (2).

$$L \stackrel{\triangle}{=} \{follow, retweet, mention, reply\} \tag{2}$$

Each layer is formed by the edges of a single label, therefore allowing in total $|L|$ layers. Each of them represents different account interaction patterns.

Definition 1 takes into consideration structural and functional graph characteristics. The former rely on combinatorial properties, whereas the latter depend directly on the Twitter activity. For the purposes of the analysis done here edges have directions, capturing the one-way nature of Twitter interaction.

Each of the Twitter graph layers is formed by a specific label and their endpoints. A given layer need not be symmetric and, in fact, such a layer is a special case. Each label in L carries a different semantic value and denotes relationships of varying strength. In the proposed approach this is translated to different edges having priority depending on their label, but this does not exclude the association of specific edge weights with different labels.

3.2. MBTI personality system

As mentioned earlier, MBTI is one of the most popular personality models stemming in large part from the theory of Jung and it aims at assessing the combination of individual cognitive functions. Specifically, in MBTI there are four independent axes, each corresponding to a function, with their two endpoints being personality attributes. This yields a total of sixteen basic personality types known by the initials of their corresponding attributes as shown in figure 1. Therein the respective frequencies are also shown. Observe that in this case as well holds the fundamental result of probability theory stating that for every discrete distribution with n data points there is at least one such point whose probability is strictly more than $1/n$. In the context of MBTI this implies there is at least one personality type which is more common than the others.

ISTJ 12%	ISFJ 8%	INFJ 4%	INTJ 7%
ISTP 4%	ISFP 3%	INFP 4%	INTP 4%
ESTP 5%	ESFP 5%	ENFP 8%	ENTP 5%
ESTJ 12%	ESFJ 8%	ENFJ 5%	ENTJ 6%

Figure 1: MBTI personalities.

The four axes representing basic cognitive functions or fundamental personality traits which determine the MBTI taxonomy are the following:

- **E**xtroversion vs **I**ntroversion: This axis determines how social an individual is. In turn this determines the nature and amount of sensory input an individual can cognitively process.
- **S**ensing vs i**N**tuition: This direction indicates the role tangible data play in a person's cognitive pro-

cess, namely whether they rely more on concrete input or abstract terms.

- **F**eeling vs **T**hinking: This variable denotes whether an individual tends to reason in order to understand a situation or relies on accumulated experience in the form of hunches.
- **P**erceiving vs **J**udging: This factor finally shows whether a person tends to place themselves inside a given situation during decision making process or they reason from a detached standpoint.

The system described above captures a major part of human decision making process as it describes how and which information is collected and furthermore how it is processed. As stated earlier the personalities of the MBTI taxonomy are shown in figure 1 in such a way that ech personality differs at exactly one trait from its neighboring ones. This property in fact extends to personalities which are adjacent if the taxonomy map would wrap. Thus, personalities are encoded in a scheme similar to that of Gray coding allowing the easy grouping by factors similar to Karnaugh maps.

In order to estimate the MBTI personality of each Twitter account in the dataset a GNN was used which operated on the entire graph. Since each of the four variables of the taxonomy are independent, their estimation was recast as four separate instances of vertex classification. The ground truth state vector contains the following attributes, which can be linked to how the basic personalities of the MBTI taxonomy manifest themselves.

- Average number of connections.
- Average number of characters.
- Average number of nonwhite characters.
- Fraction of alphabetical characters.
- Fraction of digits.
- Fraction of uppercase characters.
- Fraction of white spaces.
- Fraction of special characters.
- Average number of words.
- Fraction of unique works.
- Number of long words.
- Average word length.
- Number of unique stopwords.
- Fraction of stopwords.
- Number of sentences.
- Number of long sentences (at least 10 words).
- Average number of characters per sentence.
- Average number of words per sentence.
- Percentage of positive words.
- Percentage of negative words.
- Percentage of neutral words.

The GNN performing the vertex classification works as follows. It consists of L_0 layers where each one acts like a local spatial filter applied in parallel on the various graph segments similarly to a convolutional neural network (CNN). This is achieved through successive applications of the same nonlinear mapping $\sigma(\cdot)$, frequently reported as the activation function, on a linear transform of the graph adjacency matrix. In this equation $\mathbf{P}$ is the graph Laplacian matrix of (4), $\mathbf{H}_k$ is the output of the previous layer, and $\mathbf{W}_k$ is a trainable weight matrix.

$$\mathbf{M}_{k+1} = \sigma(\mathbf{PH}_k\mathbf{W}_k) \tag{3}$$

In (3) the graph Laplacian matrix $\mathbf{P}$ is defined as in (4), where $\mathbf{A}$ is the graph adjacency matrix and $\mathbf{D}$ is the graph *neighborhood size matrix*, namely the diagonal matrix containing the total number of labeled edges whose endpoint is the respective vertex. Since Twitter graphs are directed, then there are two adjacency matrices, one for the inbound and one for the outbound edges. Moreover, since in this work there can be multiple edges between the same vertex pair, both $\mathbf{A}$ and $\mathbf{D}$ contain the count of edges whose head or tail is the respective vertex. This implies that also the diagonal of $\mathbf{A}$ should be bolstered with the maximum number of parallel edges, since each vertex is strongly connected with itself. Instead of using an one-hot encoding for each of personality type, a vector with four entries was used as it is a more natural representation for the MBTI taxonomy.

$$\mathbf{P} \stackrel{\triangle}{=} \mathbf{D}^{-1/2}(\mathbf{A} + |L|\mathbf{I}_n)\mathbf{D}^{-1/2} \tag{4}$$

As a result of the above, there are two CNNs, one operating on inbound and another on outbound neighborhoods. The nonlinear activation function $\sigma(\cdot)$ shown in (5) is applied elementwise to the matrix argument.

$$\sigma(x) \stackrel{\triangle}{=} \ln\left(e^{\beta_0 x} + 1\right) \tag{5}$$

The synaptic weights of matrix $\mathbf{W}_k$ are also individually updated using the delta rule of equation (6). Specifically, each element of the weight matrix in the k-th layer receives a correction term of the form:

$$\Delta\mathbf{W}_k[i,j] = \eta_0\left(\frac{\beta_0 e^{\beta_0 \mathbf{M}_{k+1}[i,j]}}{1 + e^{\beta_0 \mathbf{M}_{k+1}[i,j]}}\right)\mathbf{M}_{k+1}[i,j] \tag{6}$$

The learning parameter η_0 decays with a cosine rate whose frequency ω_0 depends on the training size p_0.

When both CNN run and the personality types are obtained, then in case there is a difference between the two CNNs for the value a specific variable, it is decided by the majority of the values of the respective variables of the neighboring vertices ignoring direction. When this is not possible, then second order neighborhoods are also included, again ignoring direction.

3.3. Jump strategy

The IA proposed here moves along the edges of the graph following a heuristic mechanism for community structure discovery which eventually approximates under a set of mild assumptions a homogeneous Markov chain steady state distribution. IA starts from an arbitrary vertex and visits other vertices with probability proportional to their inbound degree locally normalized. Progressively this constructs a long sequence $\langle s_k \rangle$ containing the vertices visited by the IA, which can then be mined by any agglomerative clustering algorithm in order for the final community list to be derived. Longer sequences lead to more reliable discovery as they contain a higher variety of subsequences. Thus there are more indications whether certain vertices belong together. In particular if triangles local are successfully identified, then the more likely a community is to be properly discovered. Additional patterns assisting in community discovery are:

- **Bridges:** Losing a bridge always implies connectivity loss. Thus, two communities may be connected with at least one bridge.
- **Articulations:** They are bridge endpoints. As such, they are critical for connectivity and belong to different communities.
- **Subsequences:** Frequent subsequences are indicators that certain vertices should be grouped together, especially for shorter ones.
- **Hubs and authorities:** Both are central in communities. As such, they should be grouped with the vertices they appear with.

In the limit the IA random walk approximates the steady state distribution of a homogeneous Markov chain as the vertex selection mechanism is solely determined by the local connectivity patterns between the current and the outbound neighboring vertices. This also eliminates the effect of the starting vertex.

In order to determine the next vertex to be visited the IA makes a probabilistic decision based on a mechanism reminiscent of preferential attachment. Recall that according to the latter the probability of moving from vertex u to an outbound neighbor v is shown in equation (7). Thus, each candidate vertex v is selected with a probability proportional to its locally normalized inbound degree. The normalization constant is the sum of the in-degrees of every outbound neighbor of u.

$$\text{prob}\{u \to v\} \propto \frac{|\Gamma_i(v)|}{\sum\limits_{s \in \Gamma_o(u)} |\Gamma_i(s)|} = \frac{1}{1 + g(v)} \qquad (7)$$

In (7) the positive quantity $g(\cdot)$ which is a function of the vertex v and of the union of the inbound neighborhoods of the outbound neighborhood of the current

vertex u is obtained if both the numerator and the denominator of equation (7) are divided by $|\Gamma_i(v)|$.

$$g(v) \stackrel{\triangle}{=} \sum_{s \in \Gamma_o(u) \setminus v} \frac{|\Gamma_i(s)|}{|\Gamma_i(v)|} \qquad (8)$$

If the middle form of equation (7) is used, then $|\Gamma_i(s)|$ can be efficiently approximated under mild conditions by large set cardinality estimators [57]. Alternatively, the right hand side form of (7) is an inspiration for approximating $g(v)$ in (8) by $\hat{g}(v)$, when now the ratio of the estimators is used or, less frequently, an estimation of their ratio since ratio statistics are in general difficult to be derived. In any case, this needs to be done only once for static graphs. In that case the rightmost form of equation (7) can be approximated as in equation (9):

$$\frac{1}{1 + \hat{g}(v)} = \frac{1}{1 + e^{\ln \hat{g}(v)}} = 1 - \varphi(\ln \hat{g}(v)) \qquad (9)$$

In order to derive a probability distribution from the scores obtained by (7), the softmax map is used in this work. Specifically, the transition probability from u to an outbound neighbor v can be computed as:

$$\text{prob}\{u \to v\} = \frac{\exp(1 - \varphi(\ln \hat{g}(v)))}{\sum_s \exp(1 - \varphi(\ln \hat{g}(s)))}, \quad s \in \Gamma_o(u) \qquad (10)$$

Alternatively, given the probabilistic approximation (9), any decision rule determining the next vertex IA moves to which relies on the ratio of the transition probabilities from u to two distinct outbound neighbors v and w can use the equivalent computation of equation (11):

$$\frac{\text{prob}\{u \to v\}}{\text{prob}\{u \to w\}} = \frac{1 - \varphi(\ln \hat{g}(v))}{1 - \varphi(\ln \hat{g}(w))} \qquad (11)$$

In (9) $\varphi(\cdot)$ is the logistic function defined in equation (12). It is always positive and also it is the derivative of the softplus function commonly used as the nonlinear activation function in neural networks [58].

$$\varphi(x) \stackrel{\triangle}{=} \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}, \qquad x \in \mathbb{R} \qquad (12)$$

An important property of the logistic function which has been used in (9) is shown in (13). This property suggests a balance or a conservation law between the logistic function and its symmetric with respect to its argument and the horizontal axis $x = 1/2$.

$$\varphi(-x) = \frac{1}{1 + e^x} = \frac{e^{-x}}{1 + e^{-x}} = 1 - \varphi(x) \qquad (13)$$

Moreover, notice that $\varphi(\cdot)$ is a rescaled and shifted version of the hyperbolic tangent $\psi(\cdot)$ of (14). The latter is the Bayesian estimator of a bipolar ± 1 bit in the presence

of additive Gaussian white noise [59]. Thus the logistic function is the Bayesian estimator for a 0/1 bit.

$$\psi(x) \stackrel{\triangle}{=} \tanh(x) \stackrel{\triangle}{=} \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (14)$$

The absolute derivative value of the logistic function is bound as shown in equation (15). This means that any approximation errors of $\hat{s}$ are not magnified by $\varphi(\cdot)$.

$$\left| \frac{\partial \varphi(x)}{\partial x} \right| = \left| -\frac{-e^{-x}}{(1 + e^{-x})^2} \right| = |\varphi(x)\varphi(-x)| \leq 1 \qquad (15)$$

Extending (7) in order to include multiple labels gives (18). Therein the transition probability from u to v takes into account the parallel edges wherever they exist with additional sums raging over L. Alternatively, this can be replaced by the cardinality of the set of edges of the form $(u, v; l)$ for the various possible labels $l \in L$. Additionally, let locations of the MBTI personalities of u and v in figure 1 be (i_u, j_u) and (i_v, j_v). Then the factor $a_{u,v}$ (16) which depends on these locations is computed as follows:

$$a_{u,v} \stackrel{\triangle}{=} \exp\left(-\frac{(i_u - i_v)^2 + (j_u - j_v)^2}{2} \right) \qquad (16)$$

A last but optional modification to (7) is the addition of a forgetting factor $\lambda_{u,v}$ which penalizes outbound neighbors which are selected too frequently in favor of others. The rationale for taking into consideration past selections is that the IA should be able to escape very dense segments of a particular community in order to explore alternative paths inside the community or even move to other communities. The particular form of the forgetting factor used here is shown in equation (17). Therein $q_{u,v}$ is the number of times v has been selected as a destination for the IA from u the last γ_0 times.

$$\lambda_{u,v} \stackrel{\triangle}{=} 1 - \frac{q_{u,v}}{1 + \gamma_0} \qquad (17)$$

With these observations, the original decision rule of equation (7) is now modified to take into consideration the fact that the IA moves along a multilayer graph where vertices have their own personality according to the MBTI taxonomy giving equation (18). Moreover, the forgetting factor is also present, but its use is optional.

$$\text{prob}\{u \rightarrow v\} \propto \frac{\lambda_{u,v} a_{u,v} |\{(u, v; l)\}|}{\sum_{v \in \Gamma_o(u)} \sum_{s \in \Gamma_i(v)} |\{(s, v; l)\}|} \qquad (18)$$

The maximum number of jumps τ_0 the IA is allowed to make is determined by (19), which is a mechanism for eventually terminating the IA route. The rationale behind that limit is that the IA must visit each vertex more than once, but not too many times. Depending on

the selection of parameter ε_0 there will be sufficient visits to obtain important information regarding the role of each vertex. Because of the decision rule (18), vertices which are the endpoints of important edges will be visited more frequently, whereas vertices peripheral in large communities, with low degree, or difficult to reach will be progressively neglected up to a point.

$$\tau_0 = |V| \log^{\varepsilon_0} |V| \qquad (19)$$

Algorithm 1 IA operational framework

Require: Maximum number of hops τ_0
Ensure: Obtain vertex sequence $\langle s_k \rangle$
 1: place IA in a random vertex
 2: **repeat**
 3:    **for all** outbound vertices v **do**
 4:       compute $a_{u,v}$ from (16)
 5:    **end for**
 6:    **if** forgetting factor is enabled **then**
 7:       compute $\lambda_{u,v}$ from (17)
 8:    **end if**
 9:    compute destination s from (18)
10:    move to s **and** place s in the vertex sequence
11: **until** τ_0 is reached
12: **return** sequence $\langle s_k \rangle$

For static or slowly evolving graphs the jumps of IA can be cached such that $g(\cdot)$ can be efficiently approximated without a new aggregation of edges in the vicinity of a vertex. Exploiting this locality may considerably accelerate the computations of IA.

4. Results

4.1. Implementation

In figure 2 the components of the proposed system are shown. Therein can be seen that the IA moves on a multilayer graph stored in a standalone Neo4j instance along with the activity on it such as tweets and mentions. Another major element is the GNN which has been executed as a preprocessing step and has given to each vertex its MBTI personality. The third component is the IA itself, which is lightweght as it only has to implement the decision rule of (18) based on local input.

The operational parameters of the GNN and the IA are shown in table 2 and they pertain to various equations presented earlier in the text.

The dataset used in this work has been collected by the Twitter crawler used among others in [26]. The data therein pertains to three different graphs constructed from topic sampling using a main hashtag. Specifically, the three Twitter graphs are the following:
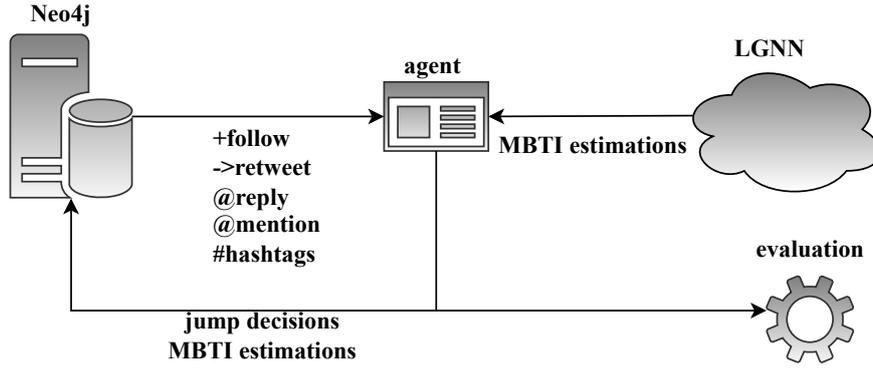
Figure 2: System architecture.

Table 2

GNN and IA Parameters

Parameter	Value
Number of layers L_0	7
Activation function $\sigma(\cdot)$	Eq. (5)
Scaling β_0	3/2
Frequency ω_0	$2\pi/(1 + p_0)$
Training size p_0	3000
Number of jumps τ_0	Eq. (19)
Factor ε_0	3/2
Decision criterion	Eq. (18)
Forgetting factor $\lambda_{u,v}$	Eq. (17)
Window size γ_0	4

- **#Julia:** This topic is about the Julia programming language. At the sampling time the next Julia developer conference was about to begin and, moreover, earlier that year a major language update was released. Therefore, there was high interest at the time for the particular topic with mostly positive or neutral feelings.
- **#Windows11:** At about the same time a major news update about the upcoming Windows 11 was made to the public. This generated mostly positive sentiment, but also some negative ones concerning the removal of some of the expected features and the question about obsolete hardware support still lingering among users.
- **#BlackList:** Just before the beginning of the final season of this the major hit it was announced that the main protagonist would not join the cast. Furthermore this was aggravated by leaked plots where her character is removed in a way deemed anticlimactic by fans. These stirred considerable controversy among them.

To form the CNN ground truth from the collected Twitter dataset two different ways were used:

- By picking the MBTI type an account has posted on their Twitter bio. This is the preferred way and it was used whenever it was possible.
- By locating a reference to an MBTI personality up to two words apart from the words *I*, *am*, *feel*, *myself*, *me*, and *being*.

Accounts with self-reported MBTI type were not included in the vertex classification by the CNN.

4.2. Evaluation

The Kullback-Leibler divergence between the distribution of the MBTI types returned by CNN for each graph compared to the global reference distribution shown in figure 1 is shown in table 4. Recall that the divergence between a distribution g and a reference one f when they are both discrete is computed by equation (20).

$$\langle g \| f \rangle \triangleq \sum_k g_k \log\left(\frac{g_k}{f_k}\right) \tag{20}$$

The low divergence between the three empirical distributions and the reference one mean that the three graphs are representative in terms of MBTI personalities.

Hashtag coherency is a functional way to assess the community structural uniformity. The rationale behind this approach is that a successful partitioning will result in communities of accounts with similar interests as expressed by hashtags. Specifically, if the Tanimoto coefficient ρ_0 between the hashtag sets of two accounts is used as a distance metric, then the minimum d_m and the average d_a intercluster distances can be used figures of merit. The results can be seen in table 5.

Table 5 should be read as follows. Three cases were tested. The first was the decision rule of (18) without the

Table 3
Twitter Social Graph Properties (from [26]).

Property	#Julia	#Win11	#BlackList	Tweets	#Julia	#Win11	#BlackList
Vertices	143019	152231	122535	Polarity% (pos/neg)	45.11/2.67	27.25/29.13	45.67/52.77
Edges	9232117	8536771	8425224	Length (avg/std)	167.33/45.12	145.17/37.83	154.86/41.84
Avg i-deg	66.21	61.89	72.43	Distinct hashtags	1182	1263	1314
Avg o-deg	71.36	63.18	76.08	Hashtags (avg/std)	5.13/0.89	8.42/1.17	7.18/1.01
Triangles	2458114	2282375	2946268	Replies (avg/std)	14.22/5.17	11.22/3.76	19.46/6.22
Squares	1034216	100736	117874	Mentions (avg/std)	17.63/4.38	13.38/3.29	15.49/5.34
Diameter	17	21	16	Density (linear/log)	64.55/1.35	56.08/1.38	68.76/1.36

Table 4
MBTI Type Distribution Divergence.

#Julia	#Win11	#BlackList
1.4413	1.2417	1.2215

Table 5
Minimum and average intercluster distances.

Metric	#Julia	#Win11	#BlackList
d_m	0.7514	0.7816	0.7932
d_a	0.7833	0.7724	0.7832
Metric	+MBTI	+MBTI	+MBTI
d_m	0.8532	0.8717	0.8365
d_a	0.8433	0.8876	0.8666
Metric	+factor	+factor	+factor
d_m	1	0.9615	0.9725
d_a	0.9918	1	0.9811

MBTI similarity factor $a_{u,v}$ and the forgetting factor $\lambda_{u,v}$. The second was the same rule with only $a_{u,v}$ and the third was with both $a_{u,v}$ and $\lambda_{u,v}$. For all three cases d_m and d_a were collected and each was normalized with respect to its respective maximum. This allows the percentage of change between jump strategies to be shown. In light of this, enabling both factors result to better communities, while excluding them both yields the worst ones.

Finally, the role of forgetting factor $\lambda_{u,v}$ it was positive as it can be seen from the above metrics. This can be attributed to the fact that, although at first iy may look counter-intuitive, it has a linear and not an exponential decay, so it forces the IA to choose less likely outbound neighbors but not too often.

5. Future Work

This conference paper focuses on the development of an intelligent agent (IA) which performs random walks on Twitter multilayer graphs with the purpose of estimating edge frequencies, which in turn heavily rely on a local decision rule for selecting the destination vertex. Moreover, destination vertices with similar MBTI profile types are given priority, whereas vertices frequently selected may be optionally penalized in order to allow the IA to escape from especially dense segments of the Twitter graph. Experimental results indicate that both factors increase the partitioning quality in terms of increased minimum and average intercluster distance.

Concerning future research directions, the most immediate one is to apply the proposed approach on more and larger benchmark Twitter graphs. Moreover, IAs may work in parallel in cooperative or adversarial modes.

Acknowledgments

References

[1] R. Calegari, G. Ciatto, V. Mascardi, A. Omicini, Logic-based technologies for multi-agent systems: A systematic literature review, Autonomous Agents and Multi-Agent Systems 35 (2021) 1–67.

[2] D. Ha, Y. Tang, Collective intelligence for deep learning: A survey of recent developments, Collective Intelligence 1 (2022).

[3] I. M. Enholm, E. Papagiannidis, P. Mikalef, J. Krogstie, Artificial intelligence and business value: A literature review, Information Systems Frontiers (2021) 1–26.

[4] S. Moussawi, M. Koufaris, R. Benbunan-Fich, How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents, Electronic Markets 31 (2021) 343–364.

[5] T. Kim, H. Song, How should intelligent agents apologize to restore trust? Interaction effects be-

tween anthropomorphism and apology attribution on trust repair, Telematics and Informatics 61 (2021).

[6] Z. Nagoev, I. Pshenokova, O. Nagoeva, Z. Sundukov, Learning algorithm for an intelligent decision making system based on multi-agent neurocognitive architectures, Cognitive Systems Research 66 (2021) 82–88.

[7] A. Gupta, S. Savarese, S. Ganguli, L. Fei-Fei, Embodied intelligence via learning and evolution, Nature communications 12 (2021) 1–12.

[8] G. Vilone, L. Longo, Notions of explainability and evaluation approaches for explainable artificial intelligence, Information Fusion 76 (2021) 89–106.

[9] K. Seaborn, N. P. Miyake, P. Pennefather, M. Otake-Matsuura, Voice in human–agent interaction: A survey, CSUR 54 (2021) 1–43.

[10] Z. Lv, D. Chen, R. Lou, A. Alazab, Artificial intelligence for securing industrial-based cyber–physical systems, Future generation computer systems 117 (2021) 291–298.

[11] E. C. Ling, I. Tussyadiah, A. Tuomi, J. Stienmetz, A. Ioannou, Factors influencing users' adoption and use of conversational agents: A systematic review, Psychology & Marketing 38 (2021) 1031–1051.

[12] S. K. Lee, P. Kavya, S. C. Lasser, Social interactions and relationships with an intelligent virtual agent, International Journal of Human-Computer Studies 150 (2021).

[13] W. Zhang, H. Liu, F. Wang, T. Xu, H. Xin, D. Dou, H. Xiong, Intelligent electric vehicle charging recommendation based on multi-agent reinforcement learning, in: The Web Conference, ACM, 2021, pp. 1856–1867.

[14] J. Lussange, I. Lazarevich, S. Bourgeois-Gironde, S. Palminteri, B. Gutkin, Modelling stock markets by multi-agent reinforcement learning, Computational Economics 57 (2021) 113–147.

[15] O. Allal-Chérif, V. Simón-Moya, A. C. C. Ballester, Intelligent purchasing: How artificial intelligence can redefine the purchasing function, Journal of Business Research 124 (2021) 69–76.

[16] S. Carta, A. Ferreira, A. S. Podda, D. R. Recupero, A. Sanna, Multi-DQN: An ensemble of deep Q-learning agents for stock market forecasting, Expert systems with applications 164 (2021).

[17] B. Wu, S. Nair, R. Martin-Martin, L. Fei-Fei, C. Finn, Greedy hierarchical variational autoencoders for large-scale video prediction, in: CVPR, IEEE/CVF, 2021, pp. 2318–2328.

[18] M. L. Olson, R. Khanna, L. Neal, F. Li, W.-K. Wong, Counterfactual state explanations for reinforcement learning agents via generative deep learning, Artificial Intelligence 295 (2021).

[19] I. H. Ahmed, C. Brewitt, I. Carlucho, F. Christianos, M. Dunion, E. Fosong, S. Garcin, S. Guo, B. Gyevnar, T. McInroe, et al., Deep reinforcement learning for multi-agent interaction, AI Communications (2022) 1–12.

[20] I.-J. Liu, U. Jain, R. A. Yeh, A. Schwing, Cooperative exploration for multi-agent deep reinforcement learning, in: ICML, PMLR, 2021, pp. 6826–6836.

[21] G. Drakopoulos, E. Kafeza, P. Mylonas, L. Iliadis, Transform-based graph topology similarity metrics, NCAA 33 (2021) 16363–16375. doi:10.1007/s00521-021-06235-9.

[22] X. Zhou, W. Liang, W. Li, K. Yan, S. Shimizu, I. Kevin, K. Wang, Hierarchical adversarial attacks against graph neural network based IoT network intrusion detection system, IEEE Internet of Things Journal (2021).

[23] F. E. Fernandes Jr, G. G. Yen, Pruning of generative adversarial neural networks for medical imaging diagnostics with evolution strategy, Information Sciences 558 (2021) 91–102.

[24] G. Drakopoulos, I. Giannoukou, S. Sioutas, P. Mylonas, Self organizing maps for cultural content delivery, NCAA 31 (2022). doi:10.1007/s00521-022-07376-1.

[25] H. Motoda, K. Saito, Critical link identification based on bridge detection for network with uncertain connectivity, in: ISMIS, Springer, 2018.

[26] G. Drakopoulos, E. Kafeza, P. Mylonas, S. Sioutas, Approximate high dimensional graph mining with matrix polar factorization: A Twitter application, in: IEEE Big Data, IEEE, 2021, pp. 4441–4449. doi:10.1109/BigData52589.2021.9671926.

[27] G. Mei, X. Wu, Y. Wang, M. Hu, J.-A. Lu, G. Chen, Compressive-sensing-based structure identification for multilayer networks, IEEE transactions on cybernetics 48 (2017) 754–764.

[28] X. Su, S. Xue, F. Liu, J. Wu, J. Yang, C. Zhou, W. Hu, C. Paris, S. Nepal, D. Jin, et al., A comprehensive survey on community detection with deep learning, IEEE Transactions on Neural Networks and Learning Systems (2022).

[29] G. Rossetti, R. Cazabet, Community discovery in dynamic networks: A survey, CSUR 51 (2018) 1–37.

[30] G. Drakopoulos, E. Kafeza, P. Mylonas, H. Al Katheeri, Building trusted startup teams from LinkedIn attributes: A higher order probabilistic analysis, in: ICTAI, IEEE, 2020, pp. 867–874. doi:10.1109/ICTAI50040.2020.00136.

[31] S. van den Beukel, S. H. Goos, J. Treur, An adaptive temporal-causal network model for social networks based on the homophily and more-becomes-more principle, Neurocomputing 338 (2019) 361–371.

[32] P. Bindu, P. S. Thilagam, D. Ahuja, Discovering suspicious behavior in multilayer social networks, Computers in Human Behavior 73 (2017) 568–582.

[33] M. Abdolali, N. Gillis, M. Rahmati, Scalable and robust sparse subspace clustering using randomized clustering and multilayer graphs, Signal Processing 163 (2019) 166–180.

[34] G. Drakopoulos, E. Kafeza, P. Mylonas, S. Sioutas, Process mining analytics for Industry 4.0 with graph signal processing, in: WEBIST, SCITEPRESS, 2021, pp. 553–560. doi:10.5220/0010718300003058.

[35] Y. Xie, M. Gong, S. Wang, B. Yu, Community discovery in networks with deep sparse filtering, Pattern Recognition 81 (2018) 50–59.

[36] X. Zheng, Z. Cai, G. Luo, L. Tian, X. Bai, Privacy-preserved community discovery in online social networks, Future Generation Computer Systems 93 (2019) 1002–1009.

[37] D. Drakopoulos, K. C. Giotopoulos, I. Giannoukou, S. Sioutas, Unsupervised discovery of semantically aware communities with tensor Kruskal decomposition: A case study in Twitter, in: SMAP, IEEE, 2020. doi:10.1109/SMAP49528.2020.9248469.

[38] T. Stahl, A. Wischnewski, J. Betz, M. Lienkamp, Multilayer graph-based trajectory planning for race vehicles in dynamic scenarios, in: ITSC, IEEE, 2019, pp. 3149–3154.

[39] V. K. Brändle, C. Galpin, H.-J. Trenz, Brexit as 'politics of division': Social media campaigning after the referendum, Social Movement Studies 21 (2022) 234–253.

[40] L. Dolega, F. Rowe, E. Branagan, Going digital? The impact of social media marketing on retail website traffic, orders and sales, Journal of Retailing and Consumer Services 60 (2021).

[41] G. Drakopoulos, A. Kanavos, I. Karydis, S. Sioutas, A. G. Vrahatis, Tensor-based semantically-aware topic clustering of biomedical documents, Computation 5 (2017). doi:10.3390/computation5030034.

[42] J. Serrano-Guerrero, F. P. Romero, J. A. Olivas, Fuzzy logic applied to opinion mining: A review, KBS 222 (2021).

[43] G. Drakopoulos, E. Kafeza, P. Mylonas, S. Sioutas, A graph neural network for fuzzy Twitter graphs, in: G. Cong, M. Ramanath (Eds.), CIKM companion volume, volume 3052, CEUR-WS.org, 2021.

[44] G. Drakopoulos, I. Giannoukou, P. Mylonas, S. Sioutas, A graph neural network for assessing the affective coherence of Twitter graphs, in: IEEE Big Data, IEEE, 2020, pp. 3618–3627. doi:10.1109/BigData50022.2020.9378492.

[45] S. Štajner, S. Yenikent, Why is MBTI personality detection from texts a difficult task?, in: European Chapter of the Association for Computational Linguistics: Main Volume, 2021, pp. 3580–3589.

[46] X. Zeng, Q. Chen, X. Fu, J. Zuo, Emotion wheel attention-based emotion distribution learning, IEEE Access 9 (2021) 153360–153370.

[47] Q. Agbaria, A. A. Mokh, Coping with stress during the coronavirus outbreak: The contribution of big five personality traits and social support, International Journal of Mental Health and Addiction 20 (2022) 1854–1872.

[48] J. Juárez, C. Santos, C. A. Brizuela, A comprehensive review and a taxonomy proposal of team formation problems, CSUR 54 (2021) 1–33.

[49] N. Cerkez, B. Vrdoljak, S. Skansi, A method for MBTI classification based on impact of class components, IEEE Access 9 (2021) 146550–146567.

[50] K. A. Nisha, U. Kulsum, S. Rahman, M. Hossain, P. Chakraborty, T. Choudhury, et al., A comparative analysis of machine learning approaches in personality prediction using MBTI, in: Computational Intelligence in Pattern Recognition, Springer, 2022, pp. 13–23.

[51] P. E. Roush, L. Atwater, Using MBTI to understand transformational leadership and self-perception accuracy, Military Psychology 4 (1992) 17–34.

[52] A. Khatua, A. Khatua, E. Cambria, Predicting political sentiments of voters from twitter in multi-party contexts, Applied Soft Computing 97 (2020).

[53] L. Lalicic, A. Huertas, A. Moreno, M. Jabreel, Emotional brand communication on Facebook and Twitter: Are DMOs successful?, Journal of Destination Marketing & Management 16 (2020).

[54] E. J. Choong, K. D. Varathan, Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in online social forum, PeerJ 9 (2021).

[55] G. Drakopoulos, I. Giannoukou, P. Mylonas, S. Sioutas, On tensor distances for self organizing maps: Clustering cognitive tasks, in: DEXA, volume 12392, Springer, 2020, pp. 195–210. doi:10.1007/978-3-030-59051-2_13.

[56] D. Fernau, S. Hillmann, N. Feldhus, T. Polzehl, Towards automated dialog personalization using MBTI personality indicators, Proc. Interspeech 2022 (2022) 1968–1972.

[57] G. Drakopoulos, S. Kontopoulos, C. Makris, Eventually consistent cardinality estimation with applications in biodata mining, in: SAC, ACM, 2016. doi:10.1145/2851613.2851887.

[58] Y. Chen, Y. Mai, J. Xiao, L. Zhang, Improving the antinoise ability of DNNs via a bio-inspired noise adaptive activation function rand softplus, Neural computation 31 (2019) 1215–1233.

[59] S. Cheng, H. Zhan, H. Yao, H. Fan, Y. Liu, Large-scale many-objective particle swarm optimizer with fast convergence based on alpha-stable mutation and logistic function, Applied Soft Computing (2020) 106947.