TOPICAL REVIEW

Review and Assessment of Digital Twin–Oriented Social Network Simulators

JIAQI WEN[iD]**, BOGDAN GABRYS**[iD]**, (Senior Member, IEEE), AND KATARZYNA MUSIAL**[iD]
Complex Adaptive Systems Laboratory, Data Science Institute, University of Technology Sydney, Sydney, NSW 2007, Australia
Corresponding author: Jiaqi Wen (Jiaqi.Wen@student.uts.edu.au)

ABSTRACT The ability to faithfully represent real social networks is critical from the perspective of testing various what-if scenarios which are not feasible to be implemented in a real system as the system's state would be irreversibly changed. High fidelity simulators allow one to investigate the consequences of different actions before introducing them to the real system. For example, in the context of social systems, an accurate social network simulator can be a powerful tool used to guide policy makers, help companies plan their advertising campaigns or authorities to analyse fake news spread. In this study we explore different Social Network Simulators (SNSs) and assess to what extent they are able to mimic the real social networks. We conduct a critical review and assessment of existing Social Network Simulators under the Digital Twin-Oriented Modelling framework proposed in our previous study. We subsequently extend one of the most promising simulators from the evaluated ones, to facilitate generation of social networks of varied structural complexity levels. This extension brings us one step closer to a Digital Twin Oriented SNS (DT Oriented SNS). We also propose an approach to assess the similarity between real and simulated networks with the composite performance indexes based on both global and local structural measures, while taking runtime of the simulator as an indicator of its efficiency. We illustrate various characteristics of the proposed DT Oriented SNS using a well known Karate Club network as an example. While not considered to be of sufficient complexity, the simulator is intended as one of the first steps on a journey towards building a Digital Twin of a social network that perfectly mimics the reality.

INDEX TERMS Social networks, network dynamics, digital twins, complex network systems.

I. INTRODUCTION

Social network simulators (SNSs) aim at faithfully representing real networked systems and attempt to model the inner rules of network dynamics. They generate simulation-based networks (built with simulated data) or hybrid networks (built with both real and simulated data) to deal with the unobservability problems resulting from data sparsity, privacy concerns and lack of ground-truth [1], [2].

As proposed in our previous study [3], Digital Twin (DT) can be seen as a modelling paradigm that serves as an accurate reflection of reality, and this means that we can treat it as an ultimate goal of the representation and modelling of any

The associate editor coordinating the review of this manuscript and approving it for publication was Yang Tang[iD].

Complex Networked System (CNS). Over the past decades, to achieve an accurate reflection of reality, variety of SNSs have been developed to embrace complexity of real world systems. Those efforts resulted in more and more realistic and complex social network simulations that, step by step, approach the desirable characteristics of a Digital Twin.

The complexity of the social network simulations results from the heterogeneity of the network components (nodes, edges and their attributes) as well as dynamically changing behaviour of those elements. Current studies have made good progress to attain a desired complexity level. They start from relatively simple simulations which are based on predetermined network statistics or connection principles about network topology, including such classical examples as Barabasi-Albert model for the scale-free network [4] or

preferential attachment of nodes based on similarity, popularity or both of them [5]. Complexity of these networks increases in the structural dimension when node attributes (e.g. age, gender, geo-space, etc.) or edge attributes (e.g. direction, weight, relationship, etc.) are introduced. To incorporate the impact of structural complexity on network simulation, social network simulators built on interaction rules have been employed due to their flexible definitions based on topology, attributes or both of them, like homophily [6], [7], triadic structure [7], [8] and geographic proximity [9].

The social network simulations get even more complex when they enable the network structure to evolve using inner rules that define how components can change over time. Studies on SNSs focus mainly on the topology change (e.g. the instantaneous social contacts [10], [11]). Only a small number of studies account for both attribute change and topology change. For example, the SNS proposed by [12] generates social networks with node attributes including features (e.g. age, gender, etc.) and preference for each feature when interacting with others, termed as social DNA (sDNA). The network topology is formed based on sDNA and evolves as sDNA changes randomly over time. However, as an example of one its existing drawbacks, the mechanisms for attributes and preferences changes need to be further studied to allow more realistic modelling.

Studies on SNSs typically focus on the social intervention analysis and conduct an optimisation of inner rules for achieving the highest possible similarity with the target network [7], [13]. The consistent network growth enabled by the iterative application of SNSs unfolds the desired network complexity at the cost of time and energy [14]. The time step in these studies is often assumed to have the same meaning with the iteration, where the networks grow linearly with time (e.g. predetermined number of added nodes or rewired edges within one iteration between time step t and $t + 1$) [7], [13]. However, the time truly spent on each iteration of SNS varies with the complexity of networks and computation resources [12] and the time that measures the duration of each SNS iteration can serve as an indicator of model efficiency. Current studies on SNSs take the similarity between simulated networks and the target networks as the objective of running an SNS [7], [12], [13], while none of them considers model efficiency.

In this study, we review and assess the current state-of-the-art SNSs under (i) the Digital Twin Oriented Modelling framework encompassing different complexity dimensions and (ii) the assessment framework based on similarity and efficiency, as proposed in our previous study [3]. To provide a possible pathway of extending SNSs towards a Digital Twin Oriented SNS, we build an inner rule-based SNS that enables the social network simulation to take into account structural and dynamic dimensions of complexity that builds on [12]. Both the similarity and the model efficiency are used as assessment criteria of the proposed SNS. We evaluate the similarity between a simulated and a real network using the composite performance indexes based on both global

and local measures and we consider runtime of SNS as an indicator of its efficiency. We use Karate Club network as an example to primarily illustrate but also examine to what extent the proposed SNS can model the structural complexity of a real social network. The diverse network patterns simulated by SNSs and their respective similarity levels to the real network assessed by the composite performance index, given different structural complexity levels, reveal the challenges of SNS performance evaluation for the future DT Oriented SNSs.

The rest of this paper is structured as follows. Section II reviews the current state-of-the-art of social network simulations and offers comparison between different SNSs in the context of modelling framework proposed in [3]. Section III presents the methodology of building an SNS to generate desired networks. Following this, section IV introduces the real data employed in the analysis and presents the results of the experiments. Finally, conclusions and future work are given in section V.

II. CURRENT STATE-OF-THE-ART OF SOCIAL NETWORK SIMULATORS

We review here and discuss the current developments in the context of SNSs from the following four perspectives: (i) modelling prerequisite – observability – that determines the observable information to be modelled and the unobservable information to be simulated; (ii) modelling generations connected with the complexity of modelling SNSs; (iii) complexity dimensions that describe the complexity of social network simulations; and (iv) assessment of SNSs concerned with the distance between the social network simulations and the target networks.

A. MODELLING PREREQUISTE

Modelling prerequisite – observability, in the context of SNSs, is concerned with the ability of reconstructing the desired complexity of networks from a limited set of observed network components in finite time with an understanding of their dynamically changing behaviour [3]. The network components include nodes, edges and attributes of nodes and edges. To generate networks with desired topology and attributes, the existing SNSs simulate the unobservable (represented with a ○) and partially observable network components (represented with a ◖) based on the observable network components (represented with a ●) and an inner rule that directs the network growth (see Table 1).

As is shown in Table 1, the simulated networks are built with purely simulated topology and attributes, dealing with the unobservability problem of all network components for the analysis of the topological features [4], [15], [16], [17], [19], [23], [24], [25], [26], [27], [28], [29], [30], [31], [33], [34], [35], [36] and mimicking of real networked systems [13], [18], [32], [53], [54].

The hybrid networks are built with the partially observable and partially simulated network components. The observable network components refer to the available information

TABLE 1. Current SNSs and the observability of network components in their simulations.

Type	Topology Nodes	Topology Edges	Attributes Nodes	Attributes Edges	Existing SNS
Simulated	○	○	○	○	[15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36],
Hybrid	●	○	●		[7], [9], [12], [37], [38], [39], [40], [41], [42]
	●	○	●	○	[43], [44], [45], [46]
	●	◗			[5], [47], [48]
	●	◗	●		[49], [50]
	●	◗		◗	[51], [52]

TABLE 2. Current SNSs and their generations.

Stage	Network	Process	Interrelation	Existing SNS
G1	△			[4], [5], [7], [13], [15], [16], [17], [18], [19] [22], [23], [24], [25], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [41], [43], [44], [45], [52],
	△	△		[54], [55], [56], [57], [58], [59]
G2a	△	▲		[60], [61]
G2b	▲			[9], [12], [20], [21], [40], [42], [46] [47], [48], [49], [50], [51], [62], [63], [64]
	▲	△		[65], [66]
G3	▲	△	△	[67]
	△	▲	△	[68], [69], [70]

related to nodes, edges and attributes, which contributes to generating realistic networks. When only partial information is observable, we simulate the unobservable network components needed for a faithful network representation. In this space, such networks, called hybrid networks, are composed of partially observable and partially simulated information related to nodes, edges and attributes. Some SNSs incorporate the observable nodes and node attributes while simulating the unobservable edges for a desired states of networks [7], [9], [12], [37], [38], [39], [42]. Some SNSs consider the partial observable edges and conduct link prediction for the unobservable edges [5], [47], [48], [49], while some SNSs incorporate edge attributes in this process [50], [51], [52]. In general, there is limited research that, in the simulation process, takes into account real-world attributes of both nodes and edges. There are some works that account for limited number of real node attributes [7], [9], [12], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [49], [50] and when it comes to edges only two simulators partially consider real-world edges [51], [52]. As a result, most of the hybrid simulators take into account some of the topology information but not many consider attribute information and this points to the need for further work on SNSs, especially in the context of learning from continuously streaming data. Such advancement would bring current SNSs closer to the Digital Twin paradigm.

B. MODELLING GENERATIONS

Modelling generations, as proposed in [3], consider three key elements of modelling complex network dynamics, the simulation of (i) networks, (ii) processes over networks, and (iii) the interrelation between the two. There are five generations of modelling paradigms (G1, G2, G3, G4 and G5), where SNSs can be built with increasing dynamics complexity through generations and finally reach the goal of a DT in G5.

We review and discuss the above mentioned three elements in the context of the modelling generations while considering two types of heterogeneity (i) the existence of a given element

(represented with a △) and in addition to its existence, (ii) the capability of an element to change over time (represented with a ▲) (See Table 2).

Generation 1 (G1) of models focuses on dynamic process on static networks (see Table 2 and Fig. 1(a)). They simulate networks that are "frozen" in time, with a dynamic process taking place on the networks where parameters of this process do not change during the simulation (e.g. epidemic spreading process on static social networks with a fixed infection rate [59], [71]). Most studies focus on SNSs in G1 [4], [5], [7], [13], [15], [16], [17], [18], [19], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [41], [43], [44], [45], [52], [54], [55], [56], [57], [58], [59], [72]. Also, many of the simulators generate static social networks without any consideration of dynamic processes [4], [5], [7], [13], [15], [16], [17], [18], [19], [22], [32], [37], [38], [39], [41], [43], [44], [45], [52].

SNSs in Generation 2 (G2) have two variations, one termed as G2a and the other as G2b. SNSs in G2a focus on evolving dynamic process on static networks (see Table 2 and Fig. 1(b)). They simulate a static network where dynamic process changes its parameters over time and gets captured in snapshots (e.g. epidemic spreading processes on static social networks with a changeable infection rate [60]). There are only few studies on SNSs in G2a [60], [61], [73].

SNSs in G2b focus on dynamic process on evolving networks (see Table 2 and Fig. 1(c)). They simulate network snapshots that describe the network topology changes over time, where the dynamic process takes place without changing its parameters (e.g. epidemic spreading process with a non-changeable infection rate on social network that evolves over time [66]). There are many studies on SNSs that model only the dynamics of network structure in this space [9], [12], [20], [21], [39], [40], [42], [46], [47], [48], [49], [50], [51], [62], [63], [64], [65], [66], [74] while few of them additionally consider the dynamic processes [65], [66].

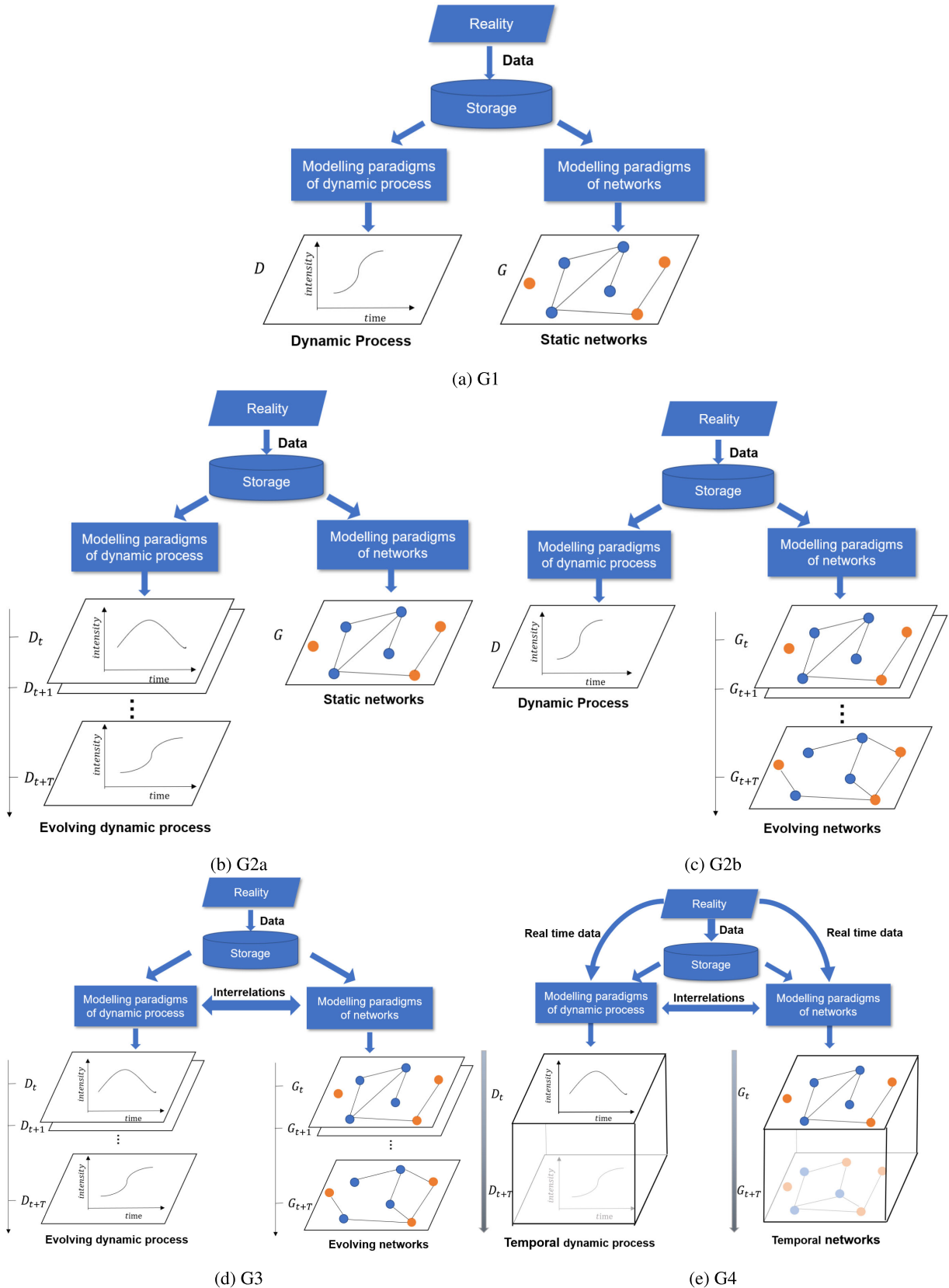SNSs in G3 focus on evolving dynamic processes on evolving networks with interrelations between them

(a) G1

(b) G2a

(c) G2b

(d) G3

(e) G4

FIGURE 1. Generations of modelling framework, including (a) G1: dynamic process on static networks, (b) G2a: evolving dynamic process on static networks, (c) G2b: dynamic process on evolving networks, (d) G3: evolving dynamic processes on evolving networks with interrelations between them, and (e) G4: temporal dynamic processes on temporal networks with interrelations between them and the continuous acquisition of real time information.
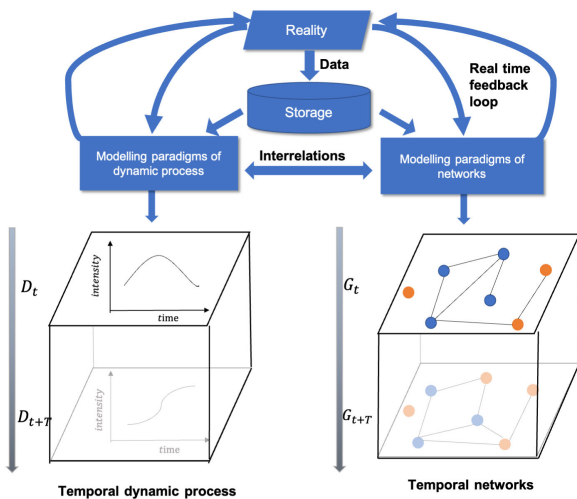
FIGURE 2. G5: the temporal dynamic processes on temporal networks, with interrelations and the real-time acquisition and feedback between the SNS and reality.

(See Fig. 1(d)). They simulate networks, dynamic processes and their changes using snapshot approach (data are processed in batches), while incorporating the interactions between a network and a process. For example, the epidemic spreading across the network can leave some nodes dead and get them removed [67], while the infection rate of epidemic spreading process can also vary depending on nodes' groups (structural patterns) [68], [69]. Only few SNSs consider the interrelations between the networks and the dynamic processes, where they just study the dynamic processes with non-changeable parameters on evolving networks [67] or evolving dynamic processes on static networks [68], [69], [70]

SNSs in G4 focus on temporal dynamic processes on temporal networks with interrelations between them and the continuous acquisition of real time information (see Fig. 1(e)). They simulate networks and the dynamic processes with instantaneous changes of network topology and parameters, while incorporating the interactions between a process and a network. SNSs in G5 further extend the SNSs in G4 by closing the feedback loop between the SNSs and the real system and can be identified as an idealised state that can be named as a Digital Twin. In G5, the temporal dynamic processes occur on temporal networks with interrelations and the real-time acquisition and feedback of information between the SNS and reality (See Fig. 2). More specifically, compared with G4, G5 is equipped with another real-time information flow from the reality to the model. This enables the two-way real-time information feedback between the SNS and reality. Currently, there are no studies on SNS in G4 and G5 and further studies are required to model such high complexity scenarios to approach a DT.

C. COMPLEXITY DIMENSIONS
Complexity dimensions describe the complexity of social networks resulting from the heterogeneity of their network

components (topology and attributes) and temporal dynamics of those components and their attributes. As proposed in [3], there are four complexity dimensions (i.e. structural, spatial, temporal and dynamics) that need to be considered when representing and modelling a network. The structural and the spatial complexities are concerned with the existence of network attributes and the space where the topology can be embedded respectively. The temporal and the dynamics complexities are connected with the changeable network components and their dynamics, respectively.

We review and discuss the network components and network dynamics of the social networks generated by the existing SNSs while considering two types of heterogeneity: (i) the existence of the components in a static time scale (represented with a △) and, in addition to its existence, (ii) the capability of the component to change over time (represented with a ▲) (see Table 3).

As is shown in Table 3, most SNSs focus on static networks with no attributes, which are categorised as G1/G2a models and characterised with the lowest level of complexity in each dimension. They generate networks based on predetermined network statistics and connection principles about topology [4], [5], [13], [15], [16], [17], [18], [19], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [41], [53], [54], [55], [56], [57]. Some other SNSs, with higher level of structural complexity, incorporate node attributes [7], [37], [38], [39], edge attributes [22], [52] or both of them [43], [44], [45] into the generation process of static networks. The SNSs in G2b/G3 have higher level of temporal complexity as network components are allowed to change over time, and they generally just consider the topology change [9], [12], [20], [40], [42], [47], [49], [50]. Few SNSs focus on changeable network dynamics, where the existing ones just consider topology and its changes over time, with a higher level of dynamics complexity but the lowest level of structural complexity [21], [48], [63]. To the best of our knowledge, there are no other SNSs in G4/G5 that would enable to model the desired level of complexity of network structure and its dynamics.

D. ASSESSMENT OF SNSs
Assessment of SNSs is concerned with investigating how close the generated networks are to the real systems that they attempt to model.

For SNSs, which generate simulated or hybrid networks to deal with the unobservable information, their assessment involves network statistics like density, degree distribution, shortest path, assortativity, modularity, clustering coefficient and betweenness [77]. Most SNSs employ network statistics concerned with the global network structures like degree centrality, betweenness centrality, and PageRank coefficient [13] to measure the similarity between the social network simulations and the target network. As an example, SNS proposed by [13] simulates the unobservable edges to achieve the desired betweenness centrality, PageRank, local clustering coefficient and degree distribution of the

TABLE 3. Current SNS and the components of their simulations.

Stage	Topology		Attributes		Network dynamics	Existing SNS
	Nodes	Edges	Nodes'	Edges'		
G1/G2a	△	△			△	[4], [15], [16], [17], [18], [13], [53], [41], [32], [19], [54], [55], [5] [56], [57], [31], [60], [59], [58], [73], [26], [23], [34], [24], [25], [27], [33], [28], [35], [36], [29], [30],
	△	△	△		△	[37], [38], [7], [39]
	△	△		△	△	[52], [22]
	△	△	△	△	△	[43], [44], [45], [75]
G2b/G3	△	▲			△	[47], [20], [76], [65]
	△	▲	△		△	[12], [9], [42], [40], [49], [50]
	△	▲	▲		△	[12], [62]
	△	▲		▲	△	[51]
	△	▲	△	▲	△	[64], [66]
	△	▲	▲	▲	△	[46]
	△	▲			▲	[48], [63], [21]

networks. There are also SNSs that simulate the missing edges for networks with partially observable edges, which can be treated as a link prediction task and evaluated considering the prediction performance with precision [49], [78], Micro Precision [79], Area Under the ROC (AUC) [78], Error Rate [80], etc. However, the above mentioned measures just focus on the states of static networks, whereas further studies are required for the measures considering network changes over time.

As an indicator of the SNSs' efficiency, the runtime of the simulation is considered by few studies [12], [62], which is proved to be influenced by the network components such as the number of simulated features and the network size.

To summarise, SNSs in the current studies can not capture enough complexity from the real world for an idealised DT modelling due to observability reasons. Though some SNSs consider real information, such as node attributes, there remains an undeveloped research space concerning how to gradually craft the increase of complexity for an improved SNS performance in one or several aspects. By increasing structural complexity, we aim to provide one of the possible pathways for extending the existing SNSs towards a DT Oriented SNS discussed in Section III. Through experiments in Section IV, we also reveal the challenges of DT Oriented modelling by presenting the complex network patterns and, thus, the difficulties they pose to the SNS performance evaluation and SNS extension towards a DT.

III. TOWARDS DIGITAL TWIN ORIENTED SOCIAL NETWORK SIMULATOR

After reviewing and evaluating current state-of-the-art SNSs in Section II, we select one of the most promising SNSs from the evaluated ones to facilitate the generations of social networks at different complexity levels based on real and simulated data. We build an inner rule-based SNS referring to and extending work in [12]. The inner rule-based SNS generate social networks based on the interaction rules that characterise nodes' preferences for connecting with others. We decide to extend this SNS various complexity levels through changes in interaction rules. We also provide an example of developing an extensible SNS towards a Digital Twin-Oriented SNS for our future study. The two steps of building such an SNS that mimics the target network include: (i) the proposal of an inner rule-based SNS with extensible complexity in section III-A and (ii) an optimised iterative application of this SNS for similarity and efficiency in section III-B considering the observability and complexity. We have included some SNSs in the Appendix A, while employing the extensible SNS initially proposed by [12] as an example in the following part of this study.

A. AN EXTENSIBLE SOCIAL NETWORK SIMULATOR

The inner rule-based SNS framework proposed by [12] models people as nodes, with the node attributes including features (i.e. such as age, gender, etc.) and individual preferences towards particular features termed as social-DNA (sDNA). The edges between these nodes represent relationships between people, which are connected according to preferential attachment rule based on topological and non-topological (attributes) characteristics. The edges can be directed or undirected and they are time-stamped with the iterative application of the SNS.

Based on this framework, we treat each node attribute as a tuple defined with a feature and sDNA (preference for

this feature and weight of preference), while allowing their variations among individuals and changes over time. This enables the social network simulations with an extensible complexity in structural, spatial, temporal and dynamics dimensions. To implement this complexity, the detailed settings about network components and network dynamics are proposed and presented below.

1) NETWORK COMPONENTS
Network components include: (i) nodes (e.g. people), (ii) edges (e.g. social contact or relations), (iii) attributes including node attributes (e.g. age, gender, geo-space, etc.) and edge attributes (e.g. direction, weight, relationship, etc.). They vary in diversity and their capability to change over time.

We assume that the SNS simulates an undirected network $\hat{G}$ with N attributed nodes within M iterations, the network formed in the k_{th} iteration is represented as

$$G_k = (V, E_k, A_k) \quad k \in [1, M] \quad (1)$$

which is composed of a fixed node set $V = \{v_1, \cdots, v_N\}$, an edge set that grows through iterations: $E_k = \{(v_i, v_j)_k|, v_i, v_j \in V, i \neq j\}$ and a node attribute set $A_k = \{a_k(v_1), \cdots, a_k(v_N)\}$ that can vary with nodes and change for each iteration.

Especially, referring to [12], the set of attributes $a_k(v_i)$ for node v_i is defined as

$$a_k(v_i) = \{f_k(v_i), p_k(v_i), w_k(v_i)\} \quad v_i \in V \quad (2)$$

which includes a feature vector $f_k(v_i)$ and the social-DNA (sDNA) defined with two vectors of the same length as $f(v_i)$. Any attribute, in this context, can be represented with a three-value tuple that is composed of feature, preference for this feature and weight of preference.

$p_k(v_i)$ determines whether to prefer similar feature with a binary value of 1 (prefer) or -1 (not prefer), and $w_k(v_i)$ represents the weight of preference for the feature with the range of $[0, 1]$.

The sDNA can be set at the level of individual nodes, groups or the whole populations in a static manner or can mutate over time. This variability of node attributes forms another source of structural complexity and requires further study.

2) NETWORK DYNAMICS
Network dynamics drives the network growth based on the current network topology and attributes. Referring to the previous study by [12], the edges of the undirected network are created without removal according to the ranking scores $s_k(v_i, v_j)$ of each node pair based on the two-way evaluation between node v_i and v_j:

$$s_k(v_i, v_j) = q_k \Phi_k(v_i, v_j) + r_k \Delta_k(v_i, v_j) + c_k^\tau \Pi_k(v_i, v_j) \quad v_i, v_j \in V, i \neq j \quad (3)$$

where $\Phi_k(v_i, v_j)$, $\Delta_k(v_i, v_j)$ and $\Pi_k(v_i, v_j)$ each represents the feature-based, popularity-based and shortest-path score

with the weights of q_k, r_k and c_k respectively. τ denotes the transposition of vector c_k in the equation.

The feature-based score is based on the nodes' preferences $sDNA(v_i)$ and $sDNA(v_j)$ for their features $f(v_i)$ and $f(v_j)$. $\odot$ is the Hadamard product, a binary operation that multiplies the corresponding elements of two matrices with the same dimension.

$$\Phi_k(v_i, v_j) = |f(v_i) - f(v_j)|^\tau (w_k(v_i) \odot l_k(v_i)) \\ + |f(v_i) - f(v_j)|^\tau (w_k(v_j) \odot l_k(v_j)) \quad (4)$$

The popularity-based score incorporates the preference for nodes with higher degrees, with m_k representing the preferential attachment parameter and d_k represents the calculation of node degrees in the k_{th} iteration:

$$\Delta_k(v_i, v_j) = m_k d_k(v_i) + m_k d_k(v_j) \quad (5)$$

The shortest path-based score is calculated based on the preference for the shortest path-length l within the range $[1, q]$ between the node pair:

$$\Pi_k(v_i, v_j) = \sum_{l=1}^{q} \gamma(v_i)^l I^l(v_i, v_j) \\ + \sum_{l=1}^{q} \gamma(v_j)^l I^l(v_i, v_j) \quad l \in [1, q] \quad (6)$$

where $\gamma(v_i)^l$ represents node v_i's weight of preference for the path-length l and $I^l(v_i, v_j)$ represents the existence of the path-length l between node pair v_i and v_j with a binary value of 0 or 1.

$$s_k(v_i, v_j) = q\Phi_k(v_i, v_j) + r\Delta_k(v_i, v_j) + c\Pi_k(v_i, v_j) \quad (7)$$

where $\Phi_k(v_i, v_j)$, $\Delta_k(v_i, v_j)$ and $\Pi_k(v_i, v_j)$ each represents the feature-based, popularity-based and shortest-path score with the weights of q, r and c.

B. OPTIMISATION TOWARDS A DIGITAL TWIN
A high quality SNS, in order to approach an idealised Digital Twin status, needs the optimised simulation of network components and network dynamics considering the three elements reviewed in section II: (i) observability, (ii) complexity, and (iii) assessment of social network simulators.

1) DEVELOPMENT PATH
Development path of the SNS towards a DT can be divided into small steps, where with each step the complexity of a simulated social network gradually increases. Under the assumption of a fixed node set and non-attributed edges, we define a development of SNS towards a DT with increasing structural complexity and temporal complexity (see Table 4).

In G1, the structural complexity increases when more features are considered with a longer feature vector $f_k(v_i)$ as part of a node attribute $a_k(v_i)$. The sDNA, p_k and w_k, represent the corresponding preference and weight of

TABLE 4. Small steps of building up the complexity.

Stage	Topology		Attributes		Dynamics
	Nodes	Edges	Nodes'	Edges'	
G1	△	△			△
	△	△	△		△
G2	△	▲	△		△
	△	▲		▲	△
G3	△	▲	▲	▲	▲

preference for respective similar features in the feature vector, composing another parts of node attribute $a_k(v_i)$. The sDNA also increases the structural complexity when more variability of the nodes' preferences (ranging from population-based, group-based to individual-based) is incorporated. Feature selection and the sDNA simulation is the necessary step required for an appropriate structural complexity level.

In G2, the temporal complexity increases when edges E_k and the node attributes A_k start to change with every simulation step, creating snapshots of evolving networks (with overall number of iterations k). This requires further study on what changes, how it changes, which network state the change leads to. In G3, the network dynamics can change with the changeable weighting factor q_k, r_k and c_k for the ranking scores of node pairs, as well as the changeable mechanism that generates sDNA through various stochastic distributions.

We can simulate the unobservable networked information and remove the unnecessary steps if the increased complexity depreciates the performance of an SNS measured using criteria of similarity and efficiency. For each step ahead, we optimise the unobservable sDNA, p_k and w_k and the other parameters with the SNS proposed in section III-A to achieve the maximised similarity (between the G_k and the desired $\hat{G}$) and the minimised runtime for its simulation with efficiency (See equation 8).

$$\min ||G_k - \hat{G}||$$
$$\min \sum_{z=1}^{k} t_z$$
$$s.t.\ p_k \in \{-1, 1\}$$
$$q_k, r_k, c_k \in (0, 1]$$
$$w_k, m_k, \gamma_k^l \in (0, 1] \tag{8}$$

The efficiency of SNS is measured with the runtime $\sum_{z=1}^{k} t_z$ spent on the k iterations of running the SNS. t_z, denoted with the index z, represents the z_{th} iteration of social network simulation. We propose the composite performance indexes to measure the similarity of social network simulations and the target network.

2) COMPOSITE PERFORMANCE INDEXES

Composite performance indexes measure the distance $||G_k - \hat{G}||$ between the simulated social network G_k and the target network $\hat{G}$ based on their similarity from both the global and the local perspectives [77]. As shown in equation 9, the composite performance index is calculated as the average distance between the simulated and the real networks in terms of each network measure η_i.

$$||G_k - \hat{G}|| = \frac{1}{n} \sum_{i=1}^{n} ||\eta_i(G_k) - \eta_i(\hat{G})|| \tag{9}$$

The network measures (node degree distribution, shortest path length distribution, modularity, assortativity, clustering coefficient distribution and the triad significance profile), each denoted by η_i, evaluate the networks from different perspectives concerning the topological characteristics. The global perspective, concerned with the high level outcomes of interactions between the number of actors, includes measures such as node degree distribution, shortest path length distribution and the values of modularity and assortativity [77]. The local perspective, connected with interpersonal relations within subgraphs, includes measures such as the clustering coefficient distribution and the triad significance profile [77], [81], [82], [83].

We employ the Jensen–Shannon divergence (JS divergence) [84] to quantify their differences in distribution and the Manhattan distance to measure the differences of normalised values. Their missing values are replaced with 1 to indicate the large gaps between the network simulations and the target network. Overall, the composite performance index of each network simulation, with values ranging from 0 to 1, is calculated as the weighted average of these values. A lower value of this quality indicator means a more similar simulation compared with the target network.

3) MULTI-OBJECTIVE OPTIMISATION

Multi-objective optimisation, towards an optimal similarity and efficiency, is required for the iterative application of DT Oriented SNS. To figure out an optimised sDNA in this process, we employ the fast elitist Non-dominated Sorting Genetic Algorithm (NSGA-2) proposed by [85] for three reasons.

First, NSGA-2 is a multi-objective evolutionary algorithm, which enables a Pareto-optimal solution set with alternative solutions in or on Pareto-optimal front [85]. Second, NSGA-2 is characterised with low computational requirements, elitist approach, and parameter-less sharing approach [85]. Third and also the most important, NSGA-2, as a type of genetic algorithm, works best in the search of multiple parameters [86] while enabling the optimisation of sDNA based on genetics selection principles. Finally, we employ the Augmented Scalarization Function (ASF) method [87] to find the best solution from the solution set obtained by NSGA-2 algorithm. It requires the definition of weights

TABLE 5. The complexity of the extended SNSs.

Extended SNSs	Stage	Topology		Attributes		Dynamics
		Nodes	Edges	Nodes'	Edges'	
Zero feature–based SNS	G1	△	△			△
Real feature–based SNS	G1	△	△	△		△
Simulated feature–based SNS	G1	△	△	△		△
hybrid feature–based SNS	G1	△	△	△		△

to characterise the importance of respective performance concerning similarity and efficiency, which enables the generation of an optimal solution.

IV. RESULTS AND ASSESSMENT

The social network simulations, which are generated using the extended SNS, vary with the node attributes. These node attributes include the feature $f_k(v_i)$ and the feature's preference: $p_k(v_i)$ and $w_k(v_i)$. Specifically, we assume a population-based preference, where all nodes share the same preference vector $p_k(v_i) \equiv p$ and $w_k(v_i) \equiv w$, without any change across iterations, which are calculated for an optimal level of similarity and efficiency based on a target network. Therefore, the extended SNSs in our experiment fall in G1 and generate static networks that are composed of fixed nodes and edges (See Table 5). The Zachary's Karate Club network [88], composed of 34 nodes with one binary attribute, is used as an illustrative example in this experiment. The nodes and edges each represent club members and their ties. Each member has a binary attribute, denoting whether they belong to the "Mr. Hi" club or the "Officer" club. For each iteration, 4% of edges are added without removal. Ten network statistics are used in the assessment stage, including the density, assortativity, modularity, shortest path length and degree distribution in the global perspective, and the clustering coefficient and the significance profile of four types of triadic closures considering the binary node attributes.

Under the above assumptions and settings, the impact of feature selection for feature vector $f(v_i)$, concerned with the structural complexity, over the iterations is studied. This involves the benchmark SNS for networks without attributes (zero feature–based SNS), SNS for networks with one real node attribute (real feature–based SNS), SNS for networks with one unobservable node attribute that is randomly simulated as 0 or 1 (simulated feature–based SNS) and SNS for networks that are built with both the real and the simulated attributes (hybrid feature–based SNS). The complexity of the generated social network and the performance of the social network simulator are each presented and analysed in section IV-B and section IV-A respectively. We also characterise the complexity of the extended SNSs in Table 5. Please note that all of these analyses are intended to provide insights into a discussion of how feasible and challenging the possible pathways to the development of more realistic DT oriented SNSs are likely to be and what one can expect along such a journey.
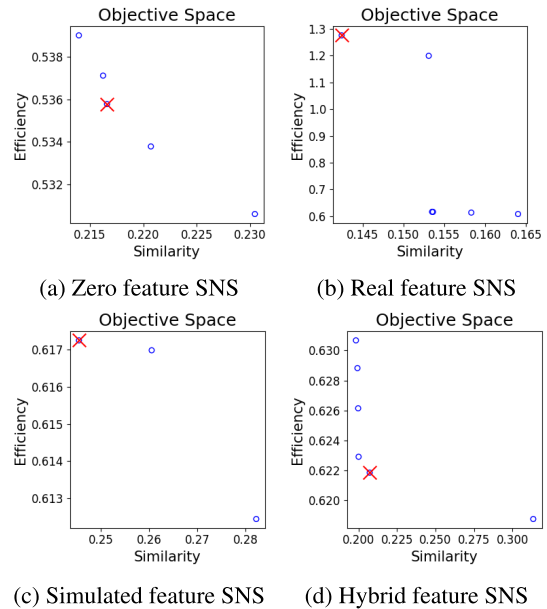

FIGURE 3. The performance of optimised SNSs based on similarity and efficiency, including zero feature–based SNS (a), real feature–based SNS (b), simulate feature–based SNS (c) and hybrid feature–based SNS (d).

A. PERFORMANCE OF THE PROPOSED SOCIAL NETWORK SIMULATOR

We measure the similarity between the simulated social networks and the target network with the composite performance indexes based on a variety of measures from both the global and the local perspectives and take the runtime of each simulation as an indicator for efficiency. Rather than proposing a performance measure or reaching final conclusions to be universally employed for SNSs, we aim to reveal the challenges of SNS performance evaluation for the future DT Oriented SNSs by discussing the diverse network patterns and the respective varying similarity levels that contribute to the composite performance index. The performance of the optimised SNSs considering multiple objectives, including similarity and efficiency, are shown in Fig. 3.

Fig. 3 presents the model performance related to similarity and efficiency for the non-discriminated solutions of SNS optimisation. For each solution, the similarity level decreases with the increase in efficiency. To find the best solution set, we employ the Augmented Scalarization Function (ASF) and attribute higher importance to the similarity objective than the efficiency objective by setting their respective weights as 0.8 (for similarity) and 0.2 (for efficiency). The best solutions, composed of the optimised parameters for the SNSs, are presented in Appendix B. Given the best solution, the performance obtained based on the similarity and efficiency is shown in Fig. 4.

As we can see in Fig. 4(a), the distance between the simulated social networks and the target network topology fluctuates over the iterations. All generated networks share a similar trend of composite performance index, which firstly decreases in the first two or three iterations and then gradually
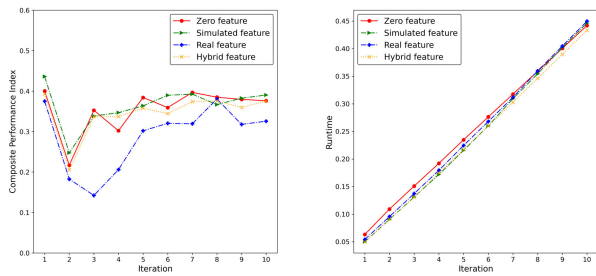
(a) The comprehensive composite performance index   (b) The runtime of the network simulation
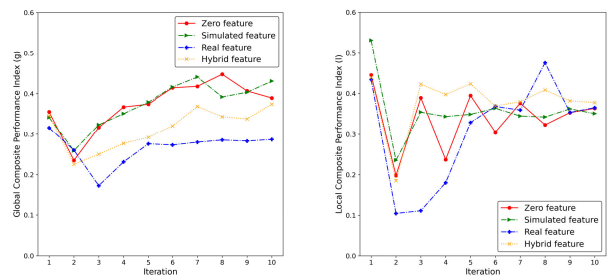
FIGURE 4. The performance of SNS based on the comprehensive composite performance index (a) and the efficiency (b).

(a) The composite performance index from a global perspective   (b) The composite performance index from a local perspective

FIGURE 5. The performance of SNS respectively based on the global and the local composite performance indexes.

increases. All the SNSs achieve the lower values of composite performance index around the second iteration. After that, the increase in composite performance index is caused by the increasing number of edges simulated in each iteration that deviates from number of edges in target network. The composite performance index first decreases and reach the lower value when network simulations approach the target with a similar number of edges. It increases when more edges are simulated, leading to deviated network patterns from the target. The real feature–based SNS, with the introduction of one feature, achieves the lowest composite performance index among all the SNSs since the second iteration, within the shortest time. The simulated feature-based SNS introduces a simulated feature to the zero feature-based SNS, which, compared with the zero feature-based and the hybrid feature-based SNSs, produces a generally higher composite performance index over the iterations.

For each SNS, the time spent on each iteration increases as more edges are added and more topological information is considered in the network growth, which, given the similar computation load of feature-based score with the non-changeable feature set, can be attributed to the calculation of popularity-based and shortest path score. There is little difference in runtime among the benchmark: zero feature–based SNS and other SNSs considering the small feature set that varies little.

Overall speaking, the introduction of the real feature to the SNS, with an increased structural complexity, improves the model performance in terms of similarity. In addition to the real feature, the consideration of the unobservable feature and its random simulation with the hybrid feature–based SNS do not bridge its distance with the target. Based on the SNS modelled with one real feature and an optimised population–based preference (the real feature-based SNS), we are able to approach the state of the target network while having the network growth captured in snapshots.

To better understand the contributions of the global measures and the local measures to the composite performance index, we respectively calculate the global composite performance index (Fig. 5(a)) and the local composite performance index (Fig. 5(b)), which each covers the five global measures and the five local measures of networks

that respectively contribute to the 50% of the composite performance index employed in the optimisation process.

Fig. 5(a) shows the values of composite performance index from the global perspective (composite performance index (g)) based on the average differences between simulations and target networks in density, modularity, assortativity, shortest path length distribution and degree distribution. For value-based measures, including density, modularity and assortativity, we use Manhattan distance to calculate their respective difference from the target. In contrast, the differences of distribution-based measures including the shortest path length and the degree distributions are measured with the JS divergence. The values of the global composite performance index (g) range from 0.17 (for real feature–based SNS) to 0.34 (for simulated feature–based SNS), and share a similar decreasing trend in the first two or three iterations with the composite performance index calculated from both the global and the local perspective (see Fig. 4). Generally all the SNS reach the lowest global composite performance index (g) around the second iteration and then deviate from the target with an increasing global composite performance index (g). The real feature-based SNS, compared with the other SNSs, achieves a lower value starting from the third iteration, similar to the trend of composite performance index calculated from both the global and the local perspective (see Fig. 4 and Fig. 5(a)).

Fig. 5(b) shows the composite performance index from the local perspective (composite performance index (l)) based on the differences between simulations and target network in the significance profiles and the clustering coefficient. Similarly, the differences in the value-based measures including the significance profiles, and the distribution-based measures including the clustering coefficient, are respectively measured with the Manhattan distance and the JS divergence. The trend of local composite performance index (l) for each SNS is similar, with a larger value range from 0.10 (for real feature–based SNS) to 0.53 (for simulated feature–based SNS), which decreases sharply in the first two iterations and then fluctuates around 0.35. Generally all the SNSs reach their lowest composite performance index (l) around 0.2, except for the real feature-based SNS, which achieves the lowest value of 0.10. The real feature-based SNS keeps a

TABLE 6. The number of nodes and edges, as well as the density of the target network.

	No. nodes	No. edges	Density	Modularity	Assortativity
Target	34	78	0.14	0.42	-0.48

lower composite performance index (l) than the other SNSs over the second, third, fourth and fifth iterations.

For each SNS, with its iterative application, the global composite performance index (g) is smaller than the local composite performance index (l) in the first iteration. This indicates that the similarity between real and simulated networks, with small number of edges added, is higher when comparing networks from the global perspective than from the local perspective. Additionally, the real feature-based SNS achieves the smallest composite performance index (best performance) with the highest level of efficiency and similarity when looking at the local and global perspectives separately as well as when they are combined.

B. COMPLEXITY OF THE SOCIAL NETWORK SIMULATION

We conduct a comparative analysis of the structural complexity built up through the iterative application of zero feature–based SNS, real feature–based SNS, simulated feature–based SNS and the hybrid feature–based SNS. The network states of social network simulations and their changes across iterations are assessed based on the global and the local measures involved in the composite performance index of the optimisation process in section III-B.

1) GLOBAL PERSPECTIVE

To have a better understanding of the network states and their changes across iterations from the global perspective connected with interactions over large number of actors, we conduct a comparative analysis of social network simulations from each SNS based on density, modularity, assortativity, degree distribution and shortest path length distribution.

2) DENSITY

Density of networks is calculated as a fraction of existing edges to all possible connections between all node pairs [77]. The density of networks increases with more edges added over the iterations. As is shown in Table 6, there are 78 edges in the target network, with a network density of 0.1390. In Fig. 6, all the simulated social networks share the same values of the edge number and density within each iteration under the assumed 4% edges (22 edges) to be added. The number of edges and the density of networks approach the values of the target network in the 4_{th} iteration.

3) MODULARITY

Modularity measures the extent to which the network is clustered and how strong those clusters are. Its value is between 0 and 1, where a larger value represents a strong community structure. As is shown in Fig. 7(a), the modularity of each
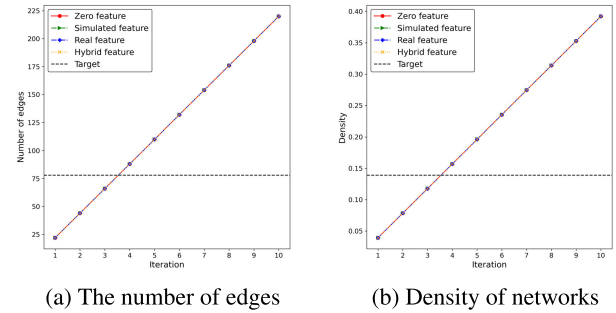

(a) The number of edges   (b) Density of networks

FIGURE 6. The number of edges and the density of social network simulations over the iterations.

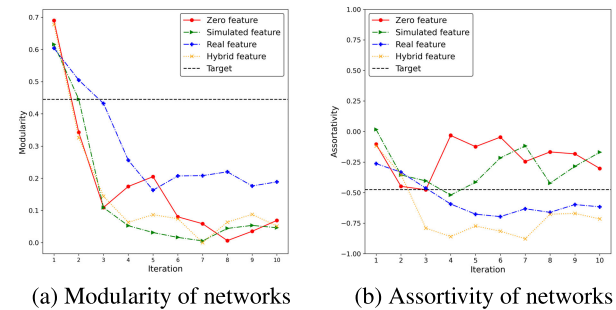(a) Modularity of networks   (b) Assortivity of networks

FIGURE 7. The modularity and the assortivity of social network simulations.

simulated network decreases over iterations, approaching the target modularity 0.4156 (See Table 6) in the second or third iteration. The real feature–based SNS generally has the largest modularity over the iterations, indicating the strongest community structure for all the simulated networks. This can be caused by the binary feature, as nodes with the same feature and the same preference tend to have similar behaviours, increasing the number of connections within a community. The social network simulated by the simulated feature-based SNS gets nearest to the target modularity in the second iteration.

4) ASSORTATIVITY

Assortativity measures whether the linked nodes have a similar degree (number of connections). Its value ranges between −1 and 1. With a larger positive value of assortativity, nodes tend to link to other nodes with the same or similar degree, indicating a stronger assortative mixing pattern [77]. As is shown in Fig. 7(b), throughout the iterations, the assortativity of simulated networks is negative and fluctuates around −0.5. This can be caused by the weak preference for higher degrees and the strong preference for dissimilar features and shortest path lengths. The zero feature-based SNS is closest to the target assortativity of −0.4762 (see Table 6) in the third iteration, while the other simulated networks approach the target in the second, third or fourth iteration.

5) DEGREE DISTRIBUTION

Degree distribution is a distribution of node degrees in a given network [77]. As shown in Table 7, the node degree of the target network fluctuates around the average value of
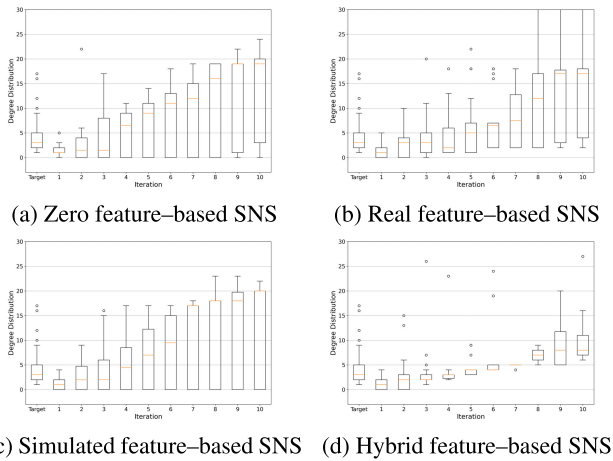
(a) Zero feature–based SNS   (b) Real feature–based SNS

(c) Simulated feature–based SNS   (d) Hybrid feature–based SNS

FIGURE 8. The degree distribution of networks.

4.59 with a standard deviation of 3.82, ranging from 1.00 to 17.00. Fig. 8 shows node degree distributions of the target network and the simulated social networks over iterations. For each SNS, the range of the degree distribution generally gets larger with more edges added over the iterations. This indicates that the node degrees vary within the network and change over time, implying the diversity of network topology. The average values of the degree distribution fluctuates around 5 with a small increase over time, which firstly approach the average node degree of the network and then deviates through iterations. The degree distributions of zero feature-based SNS (see Fig. 8(a)) and the real feature-based SNS (see Fig. 8(b)) have patterns more similar to the target network than the simulated feature-based SNS (see Fig. 8(c)) and the hybrid feature-based SNS (see Fig. 8(d)). And to quantify the level of similarity, we calculate the JS divergence based on the target degree distribution (see Fig. 9(a)).

As is shown in Fig. 9(a), all the simulated social networks have the same average node degree value, which increases systematically from 1 to 13 with the same number of edges added over the iterations. The average degree approaches the average degree of the target network in the third iteration and then deviates to higher values.

In Fig. 9(b), all the simulated networks have similar trend of JS divergence in terms of degree distribution, which firstly decreases to the lowest point at around third iteration and then increases. The hybrid feature-based SNS, in the third iteration, reaches the lowest JS divergence among all the SNS and across all iterations. Correspondingly, the degree distribution of the simulated social network obtained using the hybrid feature-based SNS in the fourth iteration (see Fig. 8(c)) has similar average values and quantiles as the target network, with few nodes having larger degree than nodes in the target network.

6) SHORTEST PATH LENGTH

Shortest path length between two nodes is defined as the number of edges along the shortest path between a pair of
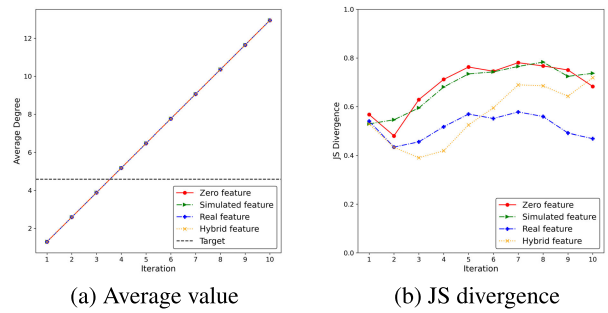

(a) Average value   (b) JS divergence

FIGURE 9. The average value and the JS divergence of degree distribution.

(a) Zero feature–based SNS   (b) Real feature–based SNS

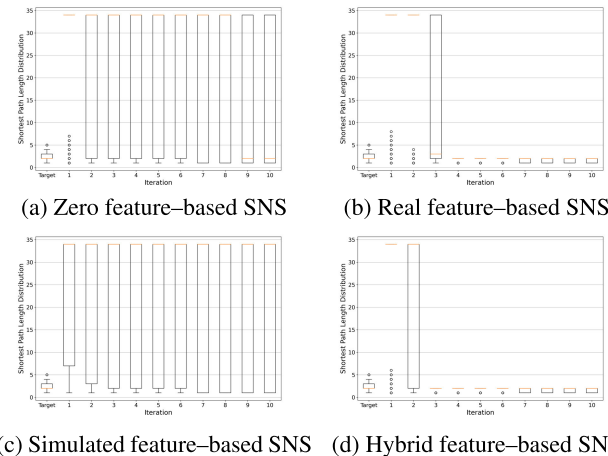(c) Simulated feature–based SNS   (d) Hybrid feature–based SNS

FIGURE 10. The shortest path length distribution of networks, where the path lengths value between 1 and 34 as we assume no self-links and replace the unavailable (infinite) paths with the lengths of 34.

nodes [77]. As shown in Table 8, the shortest path length in the target network fluctuates around an average value of 2.41 with a standard deviation of 1, ranging from 1.00 to 5.00. Fig. 10 shows the shortest path length distributions of the target network and the simulated social networks over iterations when different SNSs were used. The shortest path lengths of simulated networks value between 1 and 33 as we assume no self-links exist. The shortest path length between connected nodes can not approach an upper limit at 34 given 33 nodes. In contrast, the shortest path length between the unconnected nodes is infinite, which is hard to be measured and visualised in a distribution. Therefore, to describe the unavailable paths between unconnected node pairs, we additionally assign these node pairs with the upper limit for shortest path length 34. For real feature–based SNS and hybrid feature–based SNS, the shortest path length is distributed around the average value over 20 in the first iteration and then converges to 2. This can be caused by the unavailability of paths between node pairs given a small number of edges in the first iteration. For zero feature–based SNS and simulated feature–based SNS, the shortest path length is generally distributed around the average value over 20 over the iterations. This indicates the unavailability of paths between node pairs in the simulation process.

As is shown in Fig. 10(a), the average values of the shortest path length for all for all the social network

TABLE 7. The degree distribution of the target network.

	Average	Standard deviation	75% quantile	25% quantile	Maximum	Minimum
Target	4.59	3.82	5.00	2.00	17.00	1.00

TABLE 8. The shortest path length distribution of the target network.

	Average	Standard deviation	75% quantile	25% quantile	Maximum	Minimum
Target	2.41	0.93	3.00	2.00	5.00	1.00

(a) Average value   (b) JS divergence
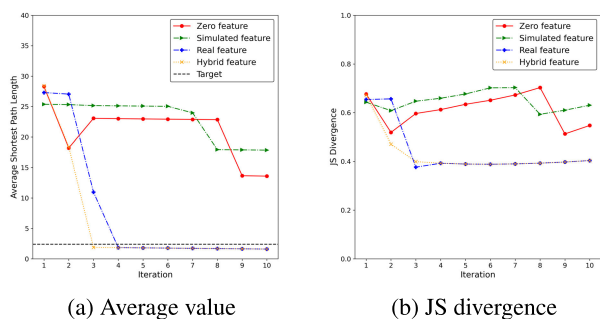
FIGURE 11. The average value and the JS divergence of shortest path length distribution.

(a) Zero feature–based SNS   (b) Real feature–based SNS

(c) Simulated feature–based SNS   (d) Hybrid feature–based SNS

FIGURE 12. The clustering coefficient distribution of network simulations.

simulations decrease with addition of edges, approach the target 2.4 and then fluctuate around 2, indicating small-world properties [77]. The hybrid feature–based SNS hits the target in the third iteration within the shortest time.

In Fig. 11(b), the JS divergence of shortest path length distribution firstly decreases to the lowest values in the third iteration and then fluctuates around 0.40 for real feature–based SNS and hybrid feature–based SNS. The real feature–based SNS and hybrid feature–based SNS gradually get all the nodes connected and finally result in the same shortest path length distributions, between the value 1 and 2, given the same number of edges (see Fig. 10). In contrast, the JS divergence of shortest path length distribution for zero feature–based SNS and simulated feature–based SNS fluctuates around 0.60. There are unavailable paths between the node paris over iterations for zero feature–based SNS and simulated feature–based SNS. The hybrid feature-based SNS achieves the lowest JS divergence in the third iteration, compared with the other SNSs across iterations.

7) LOCAL PERSPECTIVE

To have a better understanding of the network states and their changes in between, we conduct a comparative analysis of the social networks obtained from running each SNS focusing on the local perspective. We focus here on the clustering coefficient distribution and the significance profiles of the four types of triads.

8) CLUSTERING COEFFICIENT

Clustering coefficient describes the probability of a node's neighbours to be connected. Its value is between 0 and 1 [77]. As shown in Table 9, in the target network, clustering coefficient fluctuates around an average value of 0.57 with
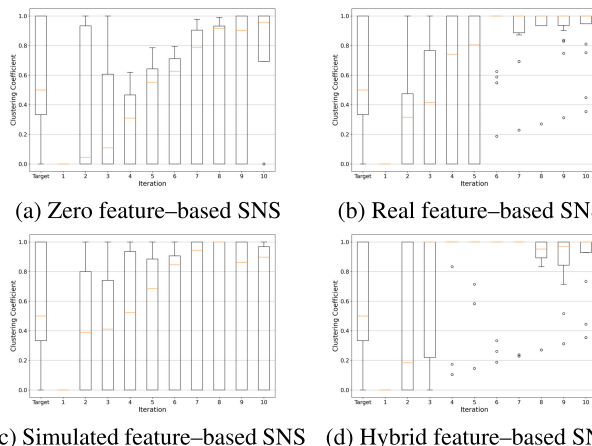
a standard deviation of 0.34, ranging from 0.00 to 1.00. Fig. 12 shows, for each SNS and each iteration, the clustering coefficient distribution of the simulated social networks. For all the networks, the values of the clustering coefficient start with 0 in the first iteration and then increase over iterations, with its distribution getting closer to that of the target and then converging to the value 1. The clustering coefficient values of the feature-based SNSs converge faster than that of the zero feature-based SNS. The real feature–based SNS gets closest to the target in the second iteration within the shortest time.

As is shown in Fig. 13(a), the average values of the clustering coefficient increase over iterations, indicating a larger probability of neighbours of one node to be connected [77]. The average clustering coefficient of the hybrid feature-based SNS converges to the higher bound 1 faster than in other SNSs and keeps steady at around 0.9 starting from the fourth iteration.

In Fig. 13(b), the JS divergences of the real feature-based SNSs firstly decrease to the lowest values at around 0.38 and then gradually increase to around 0.70. Compared with the other SNSs, the real feature-based SNS achieves the lowest JS divergence in the second generation, getting closest to the target network within the shortest time.

9) SUBGRAPHS SIGNIFICANCE PROFILE

Subgraphs significance profile describes the similarity of subgraphs and their numbers in a given network when compared to random networks of the same size and number of edges [81]. Assuming M types of subgraphs, the statistical

TABLE 9. The clustering coefficient distribution of the target network.

	Average	Standard deviation	75% quantile	25% quantile	Maximum	Minmum
Target	0.57	0.34	1.00	0.33	1.00	0.00

(a) Average value   (b) JS divergence
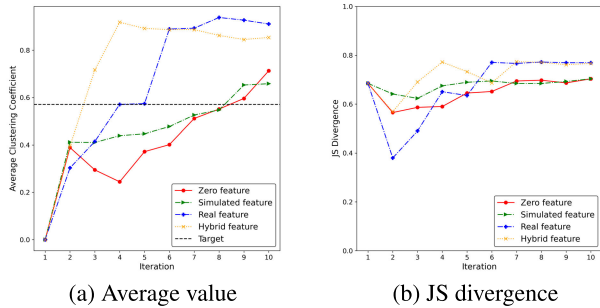
FIGURE 13. The average value and the JS divergence of clustering coefficient distribution.

Triadic closure 1   Triadic closure 2   Triadic closure 3   Triadic closure 4

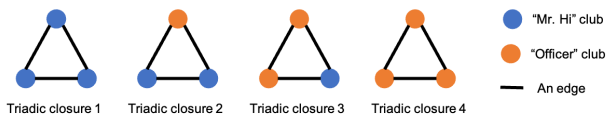● "Mr. Hi" club
● "Officer" club
— An edge

FIGURE 14. The four types of triadic closure in the network simulations.

significance of subgraph i ($i \in \{1, \cdots, M\}$) is defined by its Z-score [81], [89]:

$$Z_i = \frac{n_i - <n_i^{rand}>}{\sigma_i^{rand}} \quad i \in \{1, \cdots, M\} \quad (10)$$

where n_i, $<n_i^{rand}>$ and σ_i^{rand} represent the number of subgraph i in the studied network, the mean number of its occurrences in the random network ensemble and the corresponding standard deviation respectively. The significance profile SP_i after normalisation is defined in equation 11, where a positive value indicates a more often occurrence of subgraph i in a network than that in a set of random network ensembles.

$$SP_i = \frac{Z_i}{(\sum Z_i^2)^{1/2}} \quad (11)$$

We focus on the significance profile of a social structure, triadic closure, which is an interconnected three-node subgraph and can be categorised into four types considering the binary node attributes of the target network (See Fig. 14 and Table 10). The binary attribute determines whether a node (individual) belongs to the "Mr. Hi" club or the "Officer" club, while the edges represent the relations between the nodes. We respectively identify the triadic closures as triadic closure 1, triadic closure 2, triadic closure 3 and triadic closure 4 based on their node diversity considering the binary attribute (See Fig. 14).

In Table 10, four types of triadic closures are identified with $i \in \{1, 2, 3, 4\}$ based on a binary node attribute that decides the number of nodes attributed with the "Mr Hi" club or the "Officer" club. There are 15 observations for tragic closure 1 and 26 observations for triadic closure 4, more

TABLE 10. The definition, numbers and significance profiles of triadic closure of the target network.

Type	No. nodes		No. subgraphs	Z–score
	"Mr. Hi" club	"Officer" club		
Triadic closure 1	0	3	15	0.52
Triadic closure 2	1	2	3	-0.07
Triadic closure 3	2	1	1	-0.14
Triadic closure 4	3	0	26	0.84

(a) Triadic closure 1   (b) Triadic closure 2

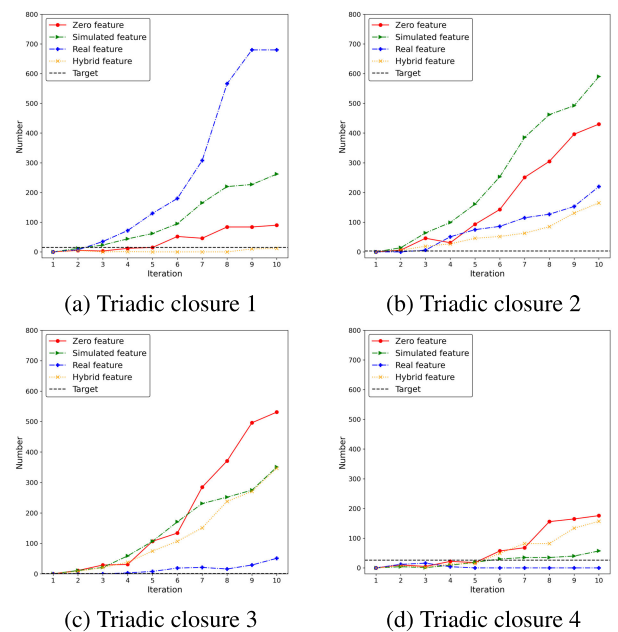(c) Triadic closure 3   (d) Triadic closure 4

FIGURE 15. The number of the four types of triadic closure in the network simulations.

than the numbers of the other two triadic closures, which means that nodes within the same club are more likely to be connected and form a triadic closure. The significance profile of triadic closure 1 and triadic closure 4 is positive, contrasting with the negative values of triadic closure 2 and triadic closure 3. The triadic closure 1 and triadic 4 appear in the target network more often than in a set of random network ensembles.

Fig. 15 shows the number of the four triadic closures in the simulated networks, which increases with addition of edges. There are more observations of triadic closure 1, 2 and 3 than triadic closure 4. All the SNSs share a similar increasing trend for triadic closure 1 and 2 over iterations, with the target number reached in around the second iteration (see Fig. 15(a) and Fig. 15(b)). The zero feature–based, the simulated feature–based and the hybrid feature–based SNSs firstly reach the target occurrences of triadic closure 3 in
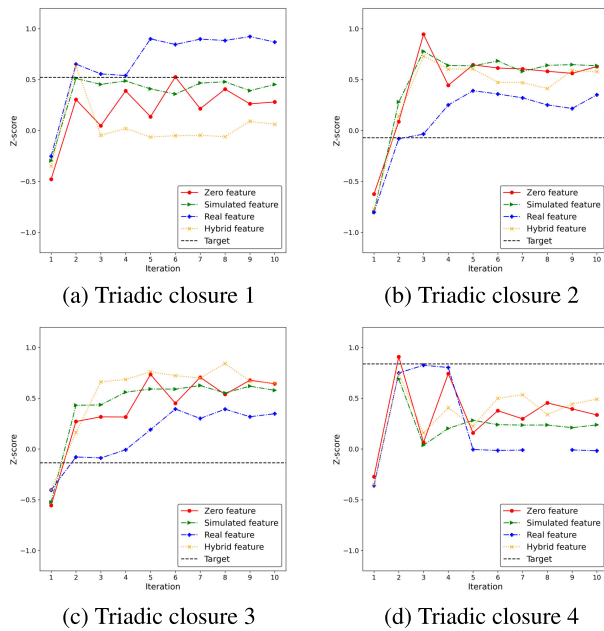
(a) Triadic closure 1

(b) Triadic closure 2

(c) Triadic closure 3

(d) Triadic closure 4

FIGURE 16. **Z–score values of triads in the social network simulations.**

the first iteration and then simulate more of them, compared with the other SNSs (see Fig. 15(c)). The zero feature-based SNS tends to have more triadic closure 4 simulated and hits the target number in the third iteration within the shortest time (see Fig. 15(d)). To better understand the significance of these triadic closures, we calculate their Z–scores (see Fig. 16).

As shown in Fig. 16, the positive/negative Z–score indicates a more/less frequent occurrence of triadic closures in comparison to random networks. There are missing Z–scores for the networks simulated by the hybrid feature-based SNS in the first two iterations. This is caused by the unavailability of triadic closures when only small number of edges are added.

The zero feature–based, the real feature–based and the simulated feature-based SNSs achieve a generally positive Z–score for triadic closure 1 over the iterations, while the hybrid feature–based SNS produces negative ones (close to zero) (see Fig. 16(a)). The SNSs achieve a positive Z–score for triadic closure 2 starting from the second iteration and afterwards deviate from the negative target (see Fig. 16(b)). All the SNSs generally enable a positive Z–score for triadic closures over iterations, except for the first iteration, where SNSs produce negative values ((see Fig. 16(a), Fig. 16(b), Fig. 16(c) and Fig. 16(d)).

The real feature–based SNS, compared with the other SNSs, attains a higher value of Z–score for the triadic closure 1, and a generally lower Z–score for the triadic closure 2, the triadic closure 3 and the triadic closure 4. The real feature–based SNS generally gets closest to the target Z–scores, which, for triadic closure 1, 2, 3 and 4, is reached in the fourth, second, second and third iteration respectively (See Fig. 16).

10) A BRIEF SUMMARY

To conclude the analysis, from a global perspective, the density, the modularity and the average node degrees increase with more edges added over the iterations for each SNS. This indicates a denser network and a stronger community structure and it is the expected result as the number of nodes remains static and only number of edges increases over the iterations. The average shortest path lengths decrease when more nodes get connected, and stabilise at a value that is smaller than 3, showing small-world properties. From a local perspective, the average clustering coefficient and number of triadic closures increase with edge addition over iterations, indicating more interpersonal connections within the subgraphs of the simulated networks.

Among all the SNSs and across iterations, the real feature–based SNS gets closest to the target when considering shortest path length distribution, clustering coefficient distribution and subgraph significance profiles. The simulated feature–based SNS gets closest to the target modularity. The zero feature–based SNS gets closest to the target assortativity. The hybrid feature–baseed SNS gets closest to the degree distribution. Generally all the network measures for all the social network simulators approach their target values and then deviate from them as the number of edges exceeds the number of edges of the target network. The SNS that achieves the closest distance to the target network measure (as measured through JS divergence) varies from case to case and requires different numbers of iterations. Real feature-based SNS can generally achieve a higher level of similarity with the target value of each measure through a smaller number of iterations within a shorter time and this indicates that including different features in the simulation process leads to different network patterns, which may respectively deviate from that of the real networks to a different extent. The trade-offs between the similarity levels considering multiple network measures pose a challenge to the performance evaluation of SNSs and their extension towards a DT. Current studies on network simulators generally employ a single network measure or combine a few to evaluate the similarity between the simulated and real social networks. The selection and combination of various features in the SNS extension and the selection and combination of multiple network measures for SNS performance evaluation in a systemic way remains a research gap and requires further study.

V. CONCLUSION

In this paper, we firstly review the current state-of-the-art of the Social Network Simulators under the modelling and assessment framework proposed in our previous study [3] and identify research gaps in the space of social network modelling. To progress the field further and achieve a simulator that can better model real networks, we extend one of the promising SNS proposed by [12] towards a Digital Twin Oriented SNS by proposing one possible pathway of increasing structural complexity, which is to include

TABLE 11. SNSs in the current studies.

SNS	Study	Link to the code	Description
VirtualSoc	[12]	https://github.com/AkandaAshraf/VirtualSoc	Dynamic Social Network Simulation Data with Ground Truth Labels and Features
Hashkat		docs.hashkat.org	A dynamic network simulor designed to model the growth of and information propagation through an online social network.
Minibaum	[23]	http://caagt.ugent.be/minibaum/	A generator for connected cubic graphs, but can be restricted to generating only graphs that have a fixed minimal girth or are bipartite
snarkhunter	[33]	http://caagt.ugent.be/cubic/	A generator for connected cubic graphs and snarks.
GenHypohamiltonian	[34]	http://caagt.ugent.be/hypoham/	A generator for hypohamiltonian and almost hypohamiltonian graphs
Genreg	[24]	http://www.mathe2.uni-bayreuth.de/markus/reggraphs.html	A very fast structure generator for regular graphs
MOLGEN	[25]	http://www.molgen.de/?src=documents/molgen4.html	A structure generator for molecular graphs.
alternating	[26]	https://github.com/nvcleemp/alternating	A generator for alternating planar graphs
CographGeneration	[27]	https://github.com/atilaajones/CographGeneration	A generator for cographs
CriticalPfreeGraph	[28]	http://caagt.ugent.be/criticalpfree/	A generator for k-critical graphs without long induced paths
nauty and Traces	[35]	https://pallini.di.uniroma1.it/	A generator for automorphism groups of graphs and digraphs. It can also produce a canonical labelling.
plantri and fullgen	[36]	https://users.cecs.anu.edu.au/ bdm/plantri/	A generator for planar graph.
Buckygen	[29]	http://caagt.ugent.be/buckygen/	A generator for all nonisomorphic fullerenes
perihamiltonian	[30]	https://github.com/nvcleemp/perihamiltonian	A generator for perihamiltonian graphs with a given connectivity.
GenerateUHG	[31]	http://caagt.ugent.be/uhg/	A generator for graphs with a given number k of hamiltonian cycles (which is especially efficient for small values of k)

informative features that are simulated or observed in the real world. With this SNS, each node attribute is composed of a node feature and a social DNA (the node's preference for a similar feature and the corresponding weight of preference).

We test different settings of sDNA in DT Oriented SNS to see which one results in social network that is closest to the target network. We propose to assess the similarity between target and simulated networks using metrics calculated at both global and local levels while taking the runtime of SNS as an indicator of efficiency. This DT Oriented SNS also serves as a tool to analyse the complexity of social network simulations built up over iterations. As an illustrative example, we conduct experiments and assessments of this DT Oriented SNS with the Karate Club network, which results show the possibility of optimisation and assessment of DT Oriented SNS towards a Digital Twin.

Literature review revealed that the majority of existing SNSs focus on generating purely simulated networks, while only a small proportion of SNSs aim to model real nodes and node attributes with/without observable information about edges. Under this constraint, we review and discuss the modelling social networks through generations towards the ultimate goal of a DT. Current SNSs generally focus on static networks or dynamic networks. Few SNSs consider networks, the dynamic process on the network and their interrelations. The interrelated social networks and dynamic processes, with the continuous acquisition of real-time information and feedback, also remain a research gap and

require further study. We also discuss the complexity of social network simulations from the perspective of four dimensions and review the current ways of assessing SNSs. Most SNSs simulate social networks composed of fixed nodes and edges, while some SNSs incorporate network attributes and slow changes in network topologies. Current SNSs consider the global level similarity between the social network simulation and the target network as an indicator of its performance, where neither the local level similarity nor the efficiency is incorporated.

In the experiment, we extend an SNS towards a DT Oriented SNS to help illustrate the possible pathways and the challenges of developing a DT Oriented SNS. The extension of an SNS towards a DT Oriented SNS incorporates different levels of structural complexity and involves experiments on the benchmark, zero feature-based SNS, and its extensions, including simulated feature-based SNS, real feature-based SNS and the hybrid feature-based SNS concerned about both the real and the simulated features. We calculate the composite performance index of different networks by combining the results of various network measures. For a deeper understanding of the composite performance index that changes over iterations, we also conduct a comparative analysis of these SNSs to see the complexity of their social network simulations over iterations. The experimental results show that among all the SNSs, the real feature-based SNS, with an appropriately increased structural complexity, has the best performance in the efficiency and similarity from both the local and the global perspective. To be more specific with

TABLE 12. Best solution for the optimisation of zero feature–based SNS considering similarity and efficiency.

Parameter	Meaning	Parameter Value			
		Zero feature SNS	Real feature SNS	Simulated feature SNS	Hybrid feature SNS
p_k	Preference vector	–	[-1]	[1]	[1,1]
w_k	Preference weighting vector	–	[0.81]	[0.17]	[0.40,0.68]
q_k	Weights of feature–based score	–	0.28	0.40	0.42
r_k	Weights of popularity–based score	0.36	0.54	0.71	0.33
m_k	Preference attachment parameter	0.04	0.06	0.25	0.46
c_k	Weights of shortest path-based score	0.30	0.83	0.88	0.80
γ^1	Weights of preference for path-length at 1	0.91	0.03	0.94	0.79
γ^2	Weights of preference for path-length at 2	0.09	0.12	0.12	0.14
γ^3	Weights of preference for path-length at 3	0.09	0.81	0.94	0.30

each measure, real feature-based SNS can also get closer to the target value through a smaller number of iterations. However, this conclusion holds for the experiments presented in this study, considering the pathway of SNS extension, the information employed in network simulation and the way we calculate the composite performance index. We need to develop a more systematic approach to developing DT Oriented SNS and propose referable SNS evaluation metrics for further study. And as SNSs add more edges over the iterations, the runtime of SNSs increases, and a stronger community structure and an assortative mixing pattern emerge, with more triadic closures as more neighbouring nodes get connected. Overall, all the measures involved in the composite performance index approach their target values and then deviate from them as the simulated network exceeds the target network's density. However, different similarity levels achieved by SNSs considering different network measures reveal the challenge of an accurate SNS performance evaluation given the specific requirements of social network simulations. This is a research gap to be addressed in our future study.

The DT Oriented SNSs can be extended with structural variations and temporal changes to approach a DT of the real systems. There is a requirement for a future study on the sDNA that varies across groups or individuals. Further research is also required on the network topology that changes over time and the process dimension that simultaneously interacts with the network dimension. Specific SNS performance evaluation criteria, considering various SNS complexity levels, also require further study. More generally, this study serves as a starting point of our future work on exploring the complexities of the real systems.

APPENDIX A
EXTENDABLE SOCIAL NETWORK SIMULATORS
In this appendix, we list some extendable social network simulators found through the review of literature and websites (http://caagt.ugent.be/CaGe/index.html; https://hog.grinvin.org/). The Table 11 below includes the names, references and links to the code for these SNSs while briefly describing their functions.

APPENDIX B
OPTIMISED PARAMETERS
In this appendix, we list the best solution for zero feature–based SNS, real feature–based SNS, simulated feature–based SNS and hybrid feature–based SNS. Table 12 below includes the names and values of the optimised parameters for SNSs.

REFERENCES
[1] H. Kavak, J.-S. Kim, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle, "Location-based social simulation," in *Proc. 16th Int. Symp. Spatial Temporal Databases*, Aug. 2019, pp. 218–221.
[2] K. Musiał and P. Kazienko, "Social networks on the Internet," *World Wide Web*, vol. 16, no. 1, pp. 31–72, 2013.
[3] J. Wen, B. Gabrys, and K. Musial, "Toward digital twin oriented modeling of complex networked systems and their dynamics: A comprehensive survey," *IEEE Access*, vol. 10, pp. 66886–66923, 2022.
[4] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[5] A. Wahid-Ul-Ashraf, M. Budka, and K. Musial, "How to predict social relationships—Physics-inspired approach to link prediction," *Phys. A, Stat. Mech. Appl.*, vol. 523, pp. 1110–1129, Jun. 2019.
[6] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociology*, vol. 27, no. 1, pp. 415–444, Aug. 2001.
[7] A. Asikainen, G. Iñiguez, J. Ureña-Carrión, K. Kaski, and M. Kivelä, "Cumulative effects of triadic closure and homophily in social networks," *Sci. Adv.*, vol. 6, no. 19, May 2020, Art. no. eaax7310.
[8] K. Musial, K. Juszczyszyn, and M. Budka, "Triad transition probabilities characterize complex networks," *Awareness Mag.*, 2012, doi: 10.2417/3201209.004369.
[9] P. Block, M. Hoffman, I. J. Raabe, J. B. Dowd, C. Rahal, R. Kashyap, and M. C. Mills, "Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world," *Nature Hum. Behav.*, vol. 4, no. 6, pp. 588–596, Jun. 2020.
[10] E. W. Fox, M. B. Short, F. P. Schoenberg, K. D. Coronges, and A. L. Bertozzi, "Modeling e-mail networks and inferring leadership using self-exciting point processes," *J. Amer. Stat. Assoc.*, vol. 111, no. 514, pp. 564–584, Apr. 2016.
[11] F. S. Passino and N. A. Heard, "Mutually exciting point process graphs for modelling dynamic networks," 2021, *arXiv:2102.06527*.
[12] A. W.-U. Ashraf, M. Budka, and K. Musial, "Simulation and augmentation of social networks for building deep learning models," 2019, *arXiv:1905.09087*.
[13] V. Arora and M. Ventresca, "Action-based modeling of complex networks," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, Jul. 2017.
[14] P. R. Hiesinger, *The Self-Assembling Brain: How Neural Networks Grow Smarter*. Princeton, NJ, USA: Princeton Univ. Press, 2021.
[15] J. P. K. Doye, "Network topology of a potential energy landscape: A static scale-free network," *Phys. Rev. Lett.*, vol. 88, no. 23, May 2002, Art. no. 238701.
[16] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[17] S. Fortunato, A. Flammini, and F. Menczer, "Scale-free network growth by ranking," *Phys. Rev. Lett.*, vol. 96, no. 21, May 2006, Art. no. 218701.

[18] V. Amati, A. Lomi, and A. Mira, "Social network modeling," *Annu. Rev. Statist. Appl.*, vol. 5, pp. 343–369, Mar. 2018.

[19] R. Solomonoff and A. Rapoport, "Connectivity of random nets," *Bull. Math. Biophys.*, vol. 13, no. 2, pp. 107–117, Jun. 1951.

[20] G. Petri and A. Barrat, "Simplicial activity driven model," *Phys. Rev. Lett.*, vol. 121, no. 22, Nov. 2018, Art. no. 228301.

[21] R. Elhesha, A. Sarkar, P. Cinaglia, C. Boucher, and T. Kahveci, "Co-evolving patterns in temporal networks of varying evolution," in *Proc. 10th ACM Int. Conf. Bioinf., Comput. Biol. Health Informat.*, Sep. 2019, pp. 494–503.

[22] Y. Lou, L. Wang, K.-F. Tsang, and G. Chen, "Towards optimal robustness of network controllability: An empirical necessary condition," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 9, pp. 3163–3174, Sep. 2020.

[23] G. Brinkmann, "Fast generation of cubic graphs," *J. Graph Theory*, vol. 23, no. 2, pp. 139–149, Oct. 1996.

[24] M. Meringer, "Fast generation of regular graphs and construction of cages," *J. Graph Theory*, vol. 30, no. 2, pp. 137–146, Feb. 1999.

[25] R. Gugisch, A. Kerber, A. Kohnert, R. Laue, M. Meringer, C. Rücker, and A. Wassermann, "MOLGEN 5.0, a molecular structure generator," in *Advances in Mathematical Chemistry and Applications*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 113–138.

[26] I. Althöfer, J. K. Haugland, K. Scherer, F. Schneider, and N. Van Cleemput, "Alternating plane graphs," *Ars Math. Contemporanea*, vol. 8, no. 2, pp. 337–363, May 2015.

[27] Á. A. Jones, F. Protti, and R. R. Del-Vecchio, "Cograph generation with linear delay," *Theor. Comput. Sci.*, vol. 713, pp. 1–10, Feb. 2018.

[28] J. Goedgebeur and O. Schaudt, "Exhaustive generation of k-critical H-free graphs," *J. Graph Theory*, vol. 87, no. 2, pp. 188–207, Feb. 2018.

[29] J. Goedgebeur and B. D. McKay, "Recursive generation of IPR fullerenes," *J. Math. Chem.*, vol. 53, no. 8, pp. 1702–1724, Sep. 2015.

[30] I. Fabrici, T. Madaras, M. Timková, N. Van Cleemput, and C. T. Zamfirescu, "Non-Hamiltonian graphs in which every edge-contracted subgraph is Hamiltonian," *Appl. Math. Comput.*, vol. 392, Mar. 2021, Art. no. 125714.

[31] J. Goedgebeur, B. Meersman, and C. T. Zamfirescu, "Graphs with few Hamiltonian cycles," *Math. Comput.*, vol. 89, no. 322, pp. 965–991, Sep. 2019.

[32] K. A. Seaton and L. M. Hackett, "Stations, trains and small-world networks," *Phys. A, Stat. Mech. Appl.*, vol. 339, nos. 3–4, pp. 635–644, Aug. 2004.

[33] G. Brinkmann and J. Goedgebeur, "Generation of cubic graphs and snarks with large girth," *J. Graph Theory*, vol. 86, no. 2, pp. 255–272, Oct. 2017.

[34] C. T. Zamfirescu and J. Goedgebeur, "On almost hypohamiltonian graphs," *Discrete Math. Theor. Comput. Sci.*, vol. 21, no. 4, pp. 1–18, 2019.

[35] B. D. McKay and A. Piperno, "Practical graph isomorphism, II," *J. Symbolic Comput.*, vol. 60, pp. 94–112, Jan. 2014.

[36] G. Brinkmann and B. D. McKay, "Fast generation of planar graphs," *MATCH Commun. Math. Comput. Chem.*, vol. 58, no. 2, pp. 323–357, 2007.

[37] X. Liu, B. Yang, W. Song, K. Musial, W. Zuo, H. Chen, and H. Yin, "A block-based generative model for attributed network embedding," *World Wide Web*, vol. 24, no. 5, pp. 1439–1464, Sep. 2021.

[38] D. R. Hunter, S. M. Goodreau, and M. S. Handcock, "Goodness of fit of social network models," *J. Amer. Stat. Assoc.*, vol. 103, no. 481, pp. 248–258, Mar. 2008.

[39] D. Verhoeven, K. Musial, S. Palmer, S. Taylor, S. Abidi, V. Zemaityte, and L. Simpson, "Controlling for openness in the male-dominated collaborative networks of the global film industry," *PLoS One*, vol. 15, no. 6, Jun. 2020, Art. no. e0234460.

[40] L. Shi, L. Zhang, and Y. Lu, "Evaluating social network-based weight loss interventions in Chinese population: An agent-based simulation," *PLoS One*, vol. 15, no. 8, Aug. 2020, Art. no. e0236716.

[41] L. Lü, C.-H. Jin, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Phys. Rev. E, Stat. Phys. Fluids Relat. Interdiscip. Top.*, vol. 80, no. 4, Oct. 2009, Art. no. 046122.

[42] Z. Boda, T. Elmer, A. Vörös, and C. Stadtfeld, "Short-term and long-term effects of a social network intervention on friendships among university students," *Sci. Rep.*, vol. 10, no. 1, p. 2889, Feb. 2020.

[43] A. Lancichinetti and S. Fortunato, "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 1, Jul. 2009, Art. no. 016118.

[44] X. Liu, W. Song, K. Musial, X. Zhao, W. Zuo, and B. Yang, "Semi-supervised stochastic blockmodel for structure analysis of signed networks," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105714.

[45] Y. Wang, D. Jin, K. Musial, and J. Dang, "Community detection in social networks considering topic correlations," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2019, vol. 33, no. 1, pp. 321–328.

[46] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "FacetNet: A framework for analyzing communities and their evolutions in dynamic networks," in *Proc. 17th Int. Conf. World Wide Web*, Apr. 2008, pp. 685–694.

[47] İ. Güneş, Ş. Gündüz-Öğüdücü, and Z. Çataltepe, "Link prediction using time series of neighborhood-based node similarity scores," *Data Mining Knowl. Discovery*, vol. 30, no. 1, pp. 147–180, Jan. 2016.

[48] F. Gao and K. Musial-Gabrys, "Hybrid structure-based link prediction model," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 1221–1228.

[49] F. Gao and K. Musial, and B. Gabrys, "A community bridge boosting social network link prediction model," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 683–689.

[50] C. Wang, V. Satuluri, and S. Parthasarathy, "Local probabilistic models for link prediction," in *Proc. 7th IEEE Int. Conf. Data Mining (ICDM)*, Oct. 2007, pp. 322–331.

[51] H. Chen and J. Li, "Exploiting structural and temporal evolution in dynamic link prediction," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 427–436.

[52] C. Dai, L. Chen, B. Li, and Y. Li, "Link prediction in multi-relational networks based on relational similarity," *Inf. Sci.*, vols. 394–395, pp. 198–216, Jul. 2017.

[53] J. Zhang, J. Wen, and J. Chen, "Modeling market fluctuations under investor sentiment with a Hawkes-contact process," *Eur. J. Finance*, vol. 29, no. 1, pp. 17–32, 2021.

[54] J. Zhang, Y. Xu, and D. Houser, "Vulnerability of scale-free cryptocurrency networks to double-spending attacks," *Eur. J. Finance*, vol. 27, no. 12, pp. 1235–1249, 2021.

[55] J. Zhang, "Influence of individual rationality on continuous double auction markets with networked traders," *Phys. A, Stat. Mech. Appl.*, vol. 495, pp. 353–392, Apr. 2018.

[56] W. Wang, J. Zhang, S. Zhao, and Y. Zhang, "Simulation of asset pricing in information networks," *Phys. A, Stat. Mech. Appl.*, vol. 513, pp. 620–634, Jan. 2019.

[57] W. Wang, S. Zhao, and J. Zhang, "Multi-asset pricing modeling using holding-based networks in energy markets," *Finance Res. Lett.*, vol. 46, May 2022, Art. no. 102483.

[58] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, no. 14, pp. 3200–3203, Apr. 2001.

[59] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Societies.*, Mar. 2005, pp. 1455–1466.

[60] P. Jovanovski, I. Tomovski, and L. Kocarev, "Modeling the spread of multiple contagions on multilayer networks," *Phys. A, Stat. Mech. Appl.*, vol. 563, Feb. 2021, Art. no. 125410.

[61] V. Carchiolo, A. Longheu, A. Malgeri, G. Mangioni, and M. Previti, "Mutual influence of users credibility and news spreading in online social networks," *Future Internet*, vol. 13, no. 5, p. 107, Apr. 2021.

[62] J.-S. Kim, H. Kavak, U. Manzoor, and A. Züfle, "Advancing simulation experimentation capabilities with runtime interventions," in *Proc. Spring Simulation Conf. (SpringSim)*, Apr./May 2019, pp. 1–11.

[63] L. Kendrick, K. Musial, and B. Gabrys, "Change point detection in social networks—Critical review with experiments," *Comput. Sci. Rev.*, vol. 29, pp. 1–13, Aug. 2018.

[64] M. Budka, K. Juszczyszyn, K. Musial, and A. Musial, "Molecular model of dynamic social network based on e-mail communication," *Social Netw. Anal. Mining*, vol. 3, no. 3, pp. 543–563, Sep. 2013.

[65] F. Liu, X. Li, and G. Zhu, "Using the contact network model and metropolis-hastings sampling to reconstruct the COVID-19 spread on the 'Diamond Princes'," *Sci. Bull.*, vol. 65, no. 15, pp. 1297–1305, Aug. 2020.

[66] J.-S. Kim, H. Jin, H. Kavak, O. C. Rouly, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle, "Location-based social network data generation based on patterns of life," in *Proc. 21st IEEE Int. Conf. Mobile Data Manage. (MDM)*, Jun./Jul. 2020, pp. 158–167.

[67] M. E. J. Newman, "Threshold effects for two pathogens spreading on a network," *Phys. Rev. Lett.*, vol. 95, no. 10, Sep. 2005, Art. no. 108701.

[68] I. Koprulu, Y. Kim, and N. B. Shroff, "Battle of opinions over evolving social networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 2, pp. 532–545, Apr. 2019.

[69] G. Fu, F. Chen, J. Liu, and J. Han, "Analysis of competitive information diffusion in a group-based population over social networks," *Phys. A, Stat. Mech. Appl.*, vol. 525, pp. 409–419, Jul. 2019.

[70] C. Pan, L.-X. Yang, X. Yang, Y. Wu, and Y. Y. Tang, "An effective rumor-containing strategy," *Phys. A, Stat. Mech. Appl.*, vol. 500, pp. 80–91, Jun. 2018.

[71] D. Król, D. Fay, and B. Gabryś, *Propagation Phenomena in Real World Networks*. Cham, Switzerland: Springer, 2015.

[72] P. Bródka, K. Musial, and J. Jankowski, "Interacting spreading processes in multilayer networks: A systematic review," *IEEE Access*, vol. 8, pp. 10316–10341, 2020.

[73] R. Eletreby, Y. Zhuang, K. M. Carley, O. Yağan, and H. V. Poor, "The effects of evolutionary adaptations on spreading processes in complex networks," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 11, pp. 5664–5670, Mar. 2020.

[74] K. Musial, M. Budka, and K. Juszczyszyn, "Creation and growth of online social network," *World Wide Web*, vol. 16, no. 4, pp. 421–447, Jul. 2013.

[75] M. Scatá, A. Di Stefano, P. Liò, and A. La Corte, "The impact of heterogeneity and awareness in modeling epidemic spreading on multiplex networks," *Sci. Rep.*, vol. 6, no. 1, p. 37105, Nov. 2016.

[76] P. Zhu, X. Wang, S. Li, Y. Guo, and Z. Wang, "Investigation of epidemic spreading process on multiplex networks by incorporating fatal properties," *Appl. Math. Comput.*, vol. 359, pp. 512–524, 2019.

[77] K. Musial, B. Gabrys, and M. Buczko, "What kind of network are you? Using local and global characteristics in network categorisation tasks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2013, pp. 1366–1373.

[78] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, Mar. 2011.

[79] H. Chen, H. Yin, X. Sun, T. Chen, B. Gabrys, and K. Musial, "Multi-level graph convolutional networks for cross-platform anchor link prediction," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 1503–1511.

[80] J. Chen, X. Wang, and X. Xu, "GC-LSTM: Graph convolution embedded LSTM for dynamic network link prediction," *Appl. Intell.*, vol. 52, pp. 7513–7528, Sep. 2021.

[81] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, "Superfamilies of evolved and designed networks," *Science*, vol. 303, no. 5663, pp. 1538–1542, Mar. 2004.

[82] M. Jia, B. Gabrys, and K. Musial, "Measuring quadrangle formation in complex networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 2, pp. 538–551, Mar. 2022.

[83] M. Jia, B. Gabrys, and K. Musial, "Directed closure coefficient and its patterns," *PLoS One*, vol. 16, no. 6, Jun. 2021, Art. no. e0253822.

[84] S. D. Pietra, V. D. Pietra, and J. Lafferty, "Inducing features of random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 380–393, Apr. 1997.

[85] K. Deb, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-2," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[86] A. Lambora, K. Gupta, and K. Chopra, "Genetic algorithm—A literature review," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Feb. 2019, pp. 380–384.

[87] A. P. Wierzbicki, "The use of reference objectives in multiobjective optimization," in *Proc. 3rd Conf. Hagen/Königswinter*, Aug. 1979, pp. 468–486.

[88] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropolog. Res.*, vol. 33, no. 4, pp. 452–473, Dec. 1977.

[89] K. Juszczyszyn, P. Kazienko, K. Musial, and B. Gabrys, "Temporal changes in connection patterns of an email-based social network," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Dec. 2008, pp. 9–12.

JIAQI WEN received the B.S. degree in applied statistics from the Central University of Finance and Economics (CUFE), Beijing, China, in 2019. She is currently pursuing the Ph.D. degree with the University of Technology Sydney.

She was a Research Assistant with the School of Economics and Management, Beihang University, Beijing. Her current research interest includes modeling complex networked systems in the digital twin space.

BOGDAN GABRYS (Senior Member, IEEE) received the M.Sc. degree in electronics and telecommunication from Silesian Technical University, Gliwice, Poland, in 1994, and the Ph.D. degree in computer science from Nottingham Trent University, Nottingham, U.K., in 1998.

Over the last 25 years, he has been with various universities and research and development departments of commercial institutions. He is currently a Professor of data science with the University of Technology Sydney, Sydney, NSW, Australia. He has published more than 200 research articles and chaired conferences, workshops, and special sessions. His current research interests include data science, complex adaptive systems, computational intelligence, machine learning, predictive analytics, and their diverse applications. He is a member of the IEEE Computational Intelligence Society and a fellow of the Higher Education Academy (HEA), U.K. He has been on program committees of a large number of international conferences with data science, computational intelligence, machine learning, and data mining themes. He is frequently invited to give keynote and plenary talks at international conferences and lectures at internationally leading research centers and commercial research labs.

KATARZYNA MUSIAL received the M.Sc. degree in computer science from the Wrocław University of Science and Technology (WrUST), Poland, the M.Sc. degree in software engineering from the Blekinge Institute of Technology, Sweden, in 2006, and the Ph.D. degree from WrUST, in November 2009.

In November 2009, she was a Senior Visiting Research Fellow with Bournemouth University (BU), where she has been a Lecturer of informatics, since 2010. In November 2011, she was a Lecturer of computer science with Kings College London. In September 2015, she returned to BU as a Principal Academic of computing, where she was a member of the Data Science Institute. In September 2017, she was an Associate Professor of network science with the School of Software, University of Technology Sydney, where she is currently a member of the Advanced Analytics Institute. Her current research interests include complex networked systems, analysis of their dynamics and evolution, adaptive and predictive modelings of their structure and characteristics, and adaptation mechanisms that exist within such systems.

• • •