Designing Optimal Interlink Structures for Interdependent Networks under Budget Constraints

Srinjoy Chattopadhyay* and Huaiyu Dai†
Department of Electrical and Computer Engineering, North Carolina State University, USA
Email: *schatto4@ncsu.edu, †hdai@ncsu.edu

Abstract

In this work, we focus on the problem of obtaining the optimal interlink structures, which maximizes the robustness of networks against random node failures, in a cost constrained setting. Using percolation theory based system equations, we have formulated our objective as a constrained optimization problem and designed algorithms serving two key purposes: i) obtaining the budget limits, $B_l$ and $B_u$, defined as the minimum budget guaranteeing the existence of a feasible and optimal interlink structure, respectively; and ii) obtaining interlink structures for intermediate budgets. Through these algorithms and associated simulation results, we demonstrate the importance of cost in network design. Furthermore, the designed algorithms have close to optimal performance while being much cheaper than cost agnostic network designs.

Index Terms
Optimal interdependence, budget constraints, network design.

I. INTRODUCTION

With the advancement of technology in the last decade, it has become evident that networks are rarely (if ever) isolated in nature. Interdependence of different network layers can be observed throughout the spectrum of the scientific studies from coupled transportation network (railways, roadways etc.) [1] to biological networks involving interactions between different proteins [2]. One fundamental phenomenon that separates interdependent networks from their isolated counterparts is the so-called process of cascading failure, where failure of a small fraction of nodes in one layer can propagate recursively amongst the different network layers leading to catastrophic effects on the network. The seminal paper by Buldyrev et al. [3] laid down the mathematical framework for analysis of interdependent networks. Subsequent works in this area focused on developing a better understanding of cascading failure on various network structures generalized from different perspectives like number of layers: double-layered or multiple-layered; nature of interdependence: one-to-one or one-to-many, dealing with the number of interlinks for each node; amount of interdependence: complete or partial, dealing with the fraction of interlinked nodes; and nature of attack: random or targeted attack on nodes. Later works like [4, 5] have focused on development of various heuristics for designing the interlink structures with the aim of maximizing the network robustness against node failures, while works like [6, 7] have focused on obtaining this optimal interlink structure mathematically at the cost of simplistic assumptions on the network model [8].

In this work, we have expanded the latter approach to take into account a very important factor which has largely been ignored in current literature: the cost of construction of interlinks. As physical constraints for network design are taken into consideration with the increasing ubiquity of interdependent networks, the interlink construction cost is likely to emerge as one of the prime factors in interdependent network design. Note that the notion of cost is not limited to the monetary case and can mean widely different things in different scenarios. For social networks, cost can be the likelihood of a person (node) to add a new connection to a person or product; whereas for infrastructure networks like coupled railway-roadway systems, cost can be interpreted as the geographical distance between two nodes. This indicates that the interlink construction cost may not have direct relationship with the topology of the network and can be influenced by numerous, in many cases intractable, factors. Due to this reason, we have considered the cost as a separate variable and have taken it as the input to our algorithm for designing optimal interlink structures under a given budget. The optimality of the interlink structures is defined with respect to (w.r.t.) the robustness of the network against node failures, which will be explained in detail in Section II. In particular, we focus on the following key aspects. Firstly, we analyze the problem of obtaining the two limits of the budget: i) the upper limit $B_u$, defined as the minimum budget above which an interlink structure maximizing robustness can always be obtained, and ii) the lower limit $B_l$, defined as the minimum budget required for designing an interlink structure. Algorithms for approximating these budget limits are also presented. Secondly, we have designed heuristic algorithms for obtaining the interlink structures for intermediate budgets to develop an understanding of the variation of maximum achievable network robustness at intermediate budgets. To the best of our knowledge, this is a first work to study the impact of cost constraints on the design of interdependent networks.

The remainder of this paper is organized as follows. Section II presents the system model and formulation of the system equations. In Section III, the objective of this work is formulated as an optimization problem. The algorithms for approximating the upper and lower budget limits, as discussed above, is presented in Section IV, while the algorithms for obtaining interlink
structures for intermediate budgets is included in Section V. Section VI discusses performance of these algorithms under different cost structures. Finally, we present the simulation results in Section VII and conclude and indicate possible future works in Section VIII.

II. SYSTEM MODEL

We consider an interdependent network comprising two layers \( A \) and \( B \) of size \( N \) each, where the layers are generated as random networks with small average intra-layer degrees \( \lambda_A, \lambda_B \ll N \), where \( \lambda_l \) denotes the average number of intralinks (solid lines in Fig. 1) in layer \( l \). This constraint ensures the locally tree-like property of the layers [9], which is required for the formulation of the system equations. The interlinks between the two layers are taken to be bidirectional and one-to-one. Bidirectionality implies the survival codependency; if nodes \( a \in A \) and \( b \in B \) are interlinked, failure of one results in the failure of the other. The one-to-one interdependence implies that every node \( a \in A \) is interlinked only to one node \( b \in B \).

Here we have abused the notation \( A \) and \( B \) to represent both the names of the constituent layers and also the set of nodes in each layer. The one-to-one structure is widely adopted in relevant literature as a first critical step towards understanding the operation and key phenomena of interdependent networks [3, 4, 5].

The critical phenomenon that makes study of interdependent networks interesting is the cascading failure mechanism, where failed nodes in one layer can initiate a recursive cascade of failures among the interdependent layers, catastrophically affecting the steady state size of the mutually connected component (MCC) of the network [3]. As we are considering a bidirectional one-to-one interlink structure, the MCC will comprise the same number of nodes in both layers. The fractional size of this component w.r.t. \( N \), represented henceforth by \( \psi \), is our desired metric of robustness in adherence to the popular approach in research on complex networks [1]-[10].

Next, we will obtain the system equation relating \( \psi \) to the fraction of attacked nodes \((1 - \eta)\) and the interlink (coupling between nodes in different layers) and intralink (coupling between nodes in same layer) structure of the network. Similar to most relevant works [3, 4, 5, 6], we have assumed the failure model to be random in nature, where the probability of a node to survive the initial attack is given by \( \eta \). In the interest of space, we omit the derivation of the system equations and simply state it. These equations have been derived in various forms, rooted from percolation and mean-field theory, based on classical works like [9, 10]. An overview of the derivation can be obtained from our previous work [7], whereas for a comprehensive understanding, we refer the reader to [11].

An indexing scheme for the nodes of the network is required for defining a general interlink structure. Without loss of generality, let us index the nodes in both layers of the network in the decreasing order of their intra-layer degree. Thus \( a_i \) and \( b_i \) denote the nodes with the \( i \)th highest intra-layer degree in layers \( A \) and \( B \), respectively. Let sequence \( \rho = \{\rho_1, \rho_2, \cdots, \rho_N\} \) be a permutation of the set \( \{1, 2, \cdots, N\} \). Let us assign to \( \rho_i \) the index of the node \( b \in B \) interlinked to \( a \in A \). Thus it is easy to see that \( \rho \) can represent any of the \( N! \) interlink structures and our objective in this work is to obtain the optimal \( \rho \) maximizing \( \psi \), subject to the budget constraints which will be defined later. With the above indexing and definitions, we can write the system equations as:

\[
\psi = \frac{1}{N} \sum_{i=1}^{N} a_i b_{\rho_i},
\]

\[
p_A = \frac{1}{N} \sum_{i=1}^{N} \tilde{a}_i b_{\rho_i}, \quad p_B = \frac{1}{N} \sum_{i=1}^{N} a_i \tilde{b}_{\rho_i},
\]
where the various terms can be defined as:

\[ a_i = \eta [1 - (1 - p_A)k_{l(i)}^{(A)}] , \]
\[ b_i = \eta [1 - (1 - p_B)k_{l(i)}^{(B)}] , \]
\[ \tilde{a}_i = \eta k_i^{(A)} \left[ 1 - (1 - p_A)k_i^{(A)} - 1 \right] / \lambda_A , \]
\[ \tilde{b}_i = \eta k_i^{(B)} \left[ 1 - (1 - p_B)k_i^{(B)} - 1 \right] / \lambda_B , \]

where \( k_i^{(l(i))} \) is the intra-layer degree of node \( i \) in layer \( l \). \( p_A \) and \( p_B \) in the above equations are interpreted as the edge percolation probability: the probability of an edge belonging to the MCC in the steady state [3, 11]. Note that the above terms (3)-(6) are themselves dependent on \( p_A \) and \( p_B \). Thus for a particular network topology and \( \eta, \psi \) is obtained by solving the self-consistent set of equations (2)-(6) iteratively [9, 10] and substituting the solution into (1).

Next, we will define the cost parameters for the construction of the interlink structure. To keep things as general as possible, we assume the cost of constructing an interlink between \( a_i \in A \) and \( b_j \in B \) to be \( c_{ij} \). The cost structure can be elegantly represented by the \( N \times N \) matrix \( C = [c_{ij}] \), henceforth referred to as the cost matrix. Thus the cost of constructing a particular interlink structure \( \rho \) is given by \( \sum_{\rho} C[i, \rho] \). We are interested in obtaining the optimal interlink structure \( (\rho) \) which maximizes the robustness of the network \( (\psi) \) subject to a budget constraint.

### III. Problem Formulation

In this section we will discuss the methods of solving the system equations (1)-(2) and formulate our objective as an optimization problem. Due to the self-consistent and interdependent nature of the system equations, it is not straightforward to obtain an optimal solution. In our previous work [7], we have shown (in Lemma III.1) that the maximization of \( \psi \) is achievable by simultaneous maximization of the three objective functions: i) \( \psi | p_A, p_B \), ii) \( p_{A_{n+1}} | p_{A_{n}}, p_{B_{n}} \), and iii) \( p_{B_{n+1}} | p_{A_{n}}, p_{B_{n}} \), where \( p_{A_{n}} \) is the value of \( p_A \) in the \( n \)th iteration; recall that the self-consistent (2) are solved iteratively. In the interest of space, we will focus on the first objective function in this work, which can be readily extended to the other cases. The important fact which facilitates the above mentioned extension is that all sequences: \( \{a\}, \{\tilde{a}\}, \{b\}, \{\tilde{b}\} \), where the sequence terms are defined in (3)-(6), are monotonically decreasing due to the indexing of the nodes in the decreasing order of their intra-layer degrees. Furthermore, the sequence elements are increasing w.r.t. \( p_A \) and \( p_B \), implying that a \( \rho \) maximizing \( (\psi | p_A, p_B) \) can be shown to maximize the other cases as well. We refer the reader to our previous work [7] for the details. We also reiterate another result from [7] which will be used for the development of the algorithms.

**Theorem III.1** For the one-to-one structure with complete interdependence, the optimal interlink structure is the matching of nodes by their intra-layer degrees in the monotonically increasing order, i.e. \( \rho^* = \{1, 2, \cdots, N\} \). Furthermore, the anti-monotonic arrangement \( \tilde{\rho}^* = \{N, N-1, \cdots, 1\} \) is the worst.

**Proof.** This result can be proved by the application of the rearrangement inequality on (1) and (2). \( \blacksquare \)

Note that similar results were heuristically suggested previously in works like [5]. Theorem III.1 indicates that in absence of any budget constraint, the monotonic arrangement is optimal. Here we tackle the same problem in a constrained setting. Mathematically, our problem statement can be written as:

\[
\max_{\rho} \psi \quad \text{s.t.} \quad \sum_{i=1}^{N} C[i, \rho_i] \leq B_{int},
\]

where \( \psi \) is given by (1), \( \rho \) is a permutation of the set \( \{1, 2, \cdots, N\} \), \( C \) is the cost matrix, and \( B_{int} \), referred to as the intermediate budget, is the budget constraint under which we are to design our interlink structure.

### IV. Budget Limits

A fundamental problem under the budget constrained setting is the identification of the lower \( (B_l) \) and upper \( (B_u) \) budget limits, which will be discussed in this section and algorithms for approximating them will be presented subsequently.

#### A. Lower limit of Budget

\( B_l \) can be defined as the minimum budget required for obtaining an interlink structure. \( B_l \) is thus the lowest cost which the network designer needs to spend in order to construct an interdependent network. Note that the definition of \( B_l \) is unrelated to the robustness \( (\psi) \) or the topology of the network. We are only interested in finding a minimum cost interlink structure. We need to resort to exhaustive search to obtain an exact value of \( B_l \). In the following, we present two algorithms to approximate \( B_l \) at a much lower computational cost.
1) **Naive algorithm:** As a first step, we have designed a naive algorithm which interlinks randomly chosen (without replacement) nodes to its most preferable (cheapest) un-interlinked node. The naive algorithm is one of the simplest approaches to approximate $B_1$ and, as the following discussion will reveal, is largely sub-optimal. Algorithm 1 starts from the lowest index and couples each node greedily to its most favorable uncoupled node. It is easy to see that there exist many strategies which can outperform the naive algorithm. However, the reason behind discussing it is that it has a very fast running time (close to $O(N)$) and for cases where the rows of $C$ are independent, its performance is comparable to the Stable Marriage Algorithm (SMA), discussed next, which has a much larger running time ($O(N^2)$). The performance comparison of these two algorithms will be presented in Section VI.

2) **Stable Marriage Algorithm:** Our next algorithm is based on the Nobel Prize winning work, popular in literature as the Stable Marriage Problem (or SMA) [12]. The work obtains a matching between a set of men and women, each having a preference list ranking the members of the opposite gender, such that all couplings are stable, i.e., the participants of any two couplings cannot exchange their partners to lead to a more favorable matching for both couples. Thus stable matching can be thought of as a locally optimal matching, where the optimality is defined w.r.t. exchanges between any two couplings. In matching theory, there exists works [13] which solve this problem w.r.t. global definitions of optimality as well. However these algorithms have a much higher time complexity ($O(N^4)$ for unweighted preference lists and $O(N^4 \log N)$ for weighted lists) which are impractical when applied to large networks. Thus as a first step towards understanding the impact of cost constraints on interlink design, we have applied the classical SMA to approximate $B_1$ and the corresponding $\rho$, by generating the preference lists for all nodes, based on $C$. Thus for any node $a_i \in A$ (or $b_i \in B$), the preference list is given by the sequence of indices which sorts $C[i, :]$ (or $C[:, i]$). This is achieved by using the \texttt{argsort} function defined in NumPy. Although the SMA can be easily obtained from numerous resources, we state it here for the reader’s ease. It will be shown in Section VI that these two algorithms have similar performance when the rows of $C$ are independent of each other. However when the rows of $C$ are correlated, Algorithm 2 outperforms Algorithm 1. The reason behind this can actually be inferred intuitively.

### Algorithm 1: Naive Algorithm for $B_1$

```plaintext
1: procedure NAIVE_BUDGET($C$)  
2:     $\rho \leftarrow$ empty array of length $N$  
3:     for $i \leftarrow$ randomly chosen without replacement from 1 to $N$ do  
4:         node $\leftarrow$ minimum $C[i, :]$, s.t. node $\notin \rho$  
5:         $\rho_i \leftarrow$ node  
6:     $B_1 \leftarrow \sum_{i=1}^N C[i, \rho_i]$  
7:     return $B_1, \rho$
```

### Algorithm 2: Stable Marriage Algorithm for $B_1$

```plaintext
1: procedure STABLE_BUDGET($C$)  
2:     $\rho \leftarrow$ empty array of length $N$  
3:     $l_m, l_w \leftarrow$ generate preference lists for all nodes  
4:     while $\exists$ uncoupled man $a$ do  
5:         $a$ proposes first $b$ in $l_m[a, :]$ whom he has not proposed yet  
6:     if $b$ is uncoupled then  
7:         $a$ and $b$ are coupled  
8:     else $\triangleright$ ($b, a'$) are coupled  
9:     $b$ chooses among $a$ and $a'$ from $l_w[b, :]$  
10:    $\rho \leftarrow$ all final couplings  
11:    $B_1 \leftarrow \sum_{i=1}^N C[i, \rho_i]$  
12:    return $B_1, \rho$
```

```
B. Upper limit of Budget

The idea behind $B_u$ is a bit more involved and deals with the robustness ($\psi$) along with $C$. As was discussed in Theorem III.1, the monotonic arrangement maximizes $\psi$ in absence of budget constraints. It will be revealed in Section VI that the monotonic arrangement of nodes is very costly. However a surprising fact, which was in fact one of the key motivating factors behind this work, is that we can design $\rho$’s which are much cheaper than the monotonic arrangement without any reduction of $\psi$, i.e. we can design interlink structures which are as good as the monotonic arrangement, in terms of robustness performance, at much lower costs. The intuition behind the design of such structures is that as (almost) all networks have multiple nodes admitting the same intra-layer degrees, the monotonic arrangement is not unique. We can design many $\rho$’s which are effectively monotonic w.r.t. the intra-layer degrees. However these $\rho$’s are highly variant in terms of their cost and thus it is possible to obtain $\rho$’s which are much cheaper than a blind monotonic arrangement with no loss in $\psi$.

Algorithm 3 Algorithm for $B_u$

1: procedure UPPER_BUDGET($C, G_1, G_2$)  
2: \hspace{1em} $\rho \leftarrow$ empty array of length $N$  
3: \hspace{1em} $l_m, l_w \leftarrow$ generate preference lists for all nodes  
4: \hspace{1em} $(k_A, k_B) \leftarrow (k_{\text{max}}, k_{\text{max}})$  
5: \hspace{1em} while all nodes are not coupled do  
6: \hspace{2em} $S_A \leftarrow \{a \in A | k(a) = k_A \text{ and } a \text{ is uncoupled}\}$  
7: \hspace{2em} $S_B \leftarrow \{b \in B | k(b) = k_B \text{ and } b \text{ is uncoupled}\}$  
8: \hspace{1em} $\triangleright$ k(i) is the intra-layer degree of node i  
9: \hspace{1em} (men,women) $\leftarrow$ $(S_A, S_B)$  
10: \hspace{1em} if $|S_A| > |S_B|$ then  
11: \hspace{2em} $k_B \leftarrow$ next highest degree  
12: \hspace{1em} else if $|S_A| < |S_B|$ then  
13: \hspace{2em} $k_A \leftarrow$ next highest degree  
14: \hspace{1em} else  
15: \hspace{2em} $k_A \leftarrow$ next highest degree  
16: \hspace{2em} $k_B \leftarrow$ next highest degree  
17: \hspace{1em} Perform matching between men and women  
18: \hspace{1em} $\rho \leftarrow$ all final couplings  
19: \hspace{1em} $B_u \leftarrow \sum_{i=1}^{N} C[i, \rho_i]$  
20: return $B_u, \rho$

Algorithm 3 is designed by scrambling the interlinks between different sets of nodes without causing any change in the intra-layer degrees of the nodes at both ends of the interlinks. The critical step is Step 17, which performs the actual matching between the two sets of nodes, referred to as men and women for ease of understanding. Similar to the case of $B_l$, we have designed the matching by the naive and the stable marriage algorithms. A key difference with the previous algorithms is that in this case, the sizes of the two sets being matched (men and women) are not the same. Thus at the end of each step, some nodes in the larger set would remain uncoupled; they would however be coupled to nodes of lower degree in the next step. The performance of these algorithms for different cost matrices will be presented in Section VI, whereas simulation results can be obtained from Section VII.

V. Intermediate budgets

The previous section discusses the definitions and algorithms for approximating $B_l$, the minimum budget required to generate an interlink structure, and $B_u$, the minimum budget above which we can always find an interlink structure (\rho) which performs optimally. The next logical step is to obtain design strategies under an intermediate budget $B_{\text{int}}$ such that $B_l \leq B_{\text{int}} \leq B_u$. Note that for $B_{\text{int}} > B_u$, $\rho$ corresponding to $B_u$ is optimal, whereas for $B_{\text{int}} < B_l$, we cannot obtain an interlink structure.

In this section, we will design different strategies to obtain interlink structures for intermediate budgets. It is important to note here that the strategies presented are by no means exhaustive. In fact, these represent the first steps towards a comprehensive understanding of interlink design under budget constraints. In the following, we present three different approaches to solving the above mentioned problem. The performance of these algorithms, under different cost matrix structures, will be presented in Section VI.
A. Greedy Cost Effective Solutions

In this approach, we start from the interlink structure obtained by the SMA to approximate $B_l$, and at each iteration of the algorithm, we choose the most cost effective exchange of couplings. The mathematical analysis for determining this cost effective exchange is presented in the following.

Let the interlink structure associated with $B_l$ be represented by $\rho$. We are interested in studying the variation in $\psi$ and the interlink construction cost when the coupling of the $I$th node in layer $A$ is changed from $\rho_I$ to $\rho_J$, i.e. the couplings $(I, \rho_I)$ and $(\rho_J, J)$ are interchanged, where $\rho_j$ represents the index of $a \in A$ which was coupled to $b_j \in B$ in $\rho$. Representing the robustness of the exchanged interlink structure as $\psi^*$ and using (1), we can write:

$$\psi^* = \frac{1}{N} \left( a_I b_J + a_{\rho_J} b_{\rho_I} + \sum_{i=1}^{N} a_i b_{p_i} \right). \quad (8)$$

Mathematical manipulation of (8) yields the following:

$$\psi^* - \psi = \frac{1}{N} (b_J - b_{\rho_I})(a_I - a_{\rho_J}), \quad (9)$$

where $\psi$ is the robustness of the original interlink structure. Using the monotonicity of the elements of $\{a\}$ and $\{b\}$, we can conclude that a particular exchange would lead to an increase in $\psi$ if either of the following conditions are true:

- $J > \rho_I$ and $I > \rho_J$
- $J < \rho_I$ and $I < \rho_J$

The change in interlink construction cost can be computed as follows:

$$\Delta B = C[I, J] + C[\rho_J, \rho_I] - C[I, \rho_I] - C[\rho_J, J]. \quad (10)$$

To obtain an interlink structure satisfying $B_{int}$, we have designed a greedy algorithm which chooses the most cost effective exchange of couplings at each stage utilizing (9) and (10). The cost effectiveness is defined as the exchange which leads to the maximum increase in $\psi$ ($\psi^* - \psi$) with the minimum increase in budget ($\Delta B$). In particular, we have chosen exchanges which lead to the highest value of $\Delta \psi / \Delta B$. Starting from $B_l$, the iterations are repeated until the required budget $B_{int}$ is reached.

B. Greedy Node Degree based Solutions

Another approach towards solving the interlink problem is to look at it from the point of view of intra-layer node degrees. Recall that monotonic arrangement is optimal for unconstrained budget. Thus for $B_{int}$ between $B_l$ and $B_u$, a possible strategy could be to select a fraction of nodes and construct couplings based on the monotonic arrangement for these particular nodes, while the remaining nodes are interlinked to achieve minimum cost. We experimented with several heuristics and found out that selection of nodes with the lowest degrees and coupling them monotonically, performs the best. An intuitive reasoning behind this can be obtained from the elements of the sequences given by (3)-(6). Note that with the increase of the intra-layer degrees ($k_i^{[A]}$, $k_i^{[B]}$), the elements approach 1 at an exponential rate. Thus the variation among elements corresponding to lower degrees is much larger than that for the higher degrees, where elements are close to 1. This leads to an interesting conjecture that lower degree nodes might play a more important role in determining the robustness of an interlink structure under a budget constrained setting.

C. Genetic Algorithm Based Solution

The third approach by which we have obtained interlink structures for intermediate budgets is by using the genetic algorithm. Genetic algorithms have been known to solve many tough optimization problems by mimicking the evolutionary process of living organisms. In particular, we have designed our algorithm based on works for solving the generalized and quadratic assignment problems [14], which bear a close resemblance to our problem. However following the algorithms mentioned in these works led to a major roadblock: the quality of the solution is strongly dependent on the quality of the initial population and thus random generation of initial population leads to very poor results. To circumvent this issue, we designed the generation of the initial population in a different way such that a fraction of the initial population has a good quality. The basic idea behind this is a differential smudging of the interlink structure ($\rho$) associated with $B_u$. As explained in the previous section, this $\rho$ couples nodes in the monotonic order w.r.t. the intra-layer degrees. To generate the initial population, we have randomly smudged, i.e. locally scrambled by a randomly generated scrambling length at randomly selected points, $\rho$ so as to cause slight changes in the corresponding $\psi$ and cost. We have increased the mean scrambling length until a known fraction (chosen to be 10%) of the initial population satisfies the budget constraint $B_{int}$. This smudge based population generation ensures that a fraction of genes (interlink structures) in the initial population is of good quality. Algorithm 4 incorporates a lot of ideas which we have not discussed above. We will provide a brief description of these in the following.

i) Child generation from parents: The basic idea behind this is to combine two different parent genes, $\rho_1$ and $\rho_2$, into $\rho_3$, which
Algorithm 4 Genetic Algorithm for $B_{int}$

1: procedure GENETIC_INT_BUDGET($C, G_1, G_2$) 
2: Obtain initial population based on randomized differential smudging 
3: while population converges do 
4: Select two random interlink structures $\rho_1$ and $\rho_2$ from the population. 
5: Combine $\rho_1$ and $\rho_2$ to generate a child $\rho_3$. 
6: Replace the interlink structure with the worst quality from the population with $\rho_3$. 
7: $\rho \leftarrow$ interlink structure of highest quality 
8: return $\rho$

<table>
<thead>
<tr>
<th></th>
<th>$B_{int}$</th>
<th>$B_{naive}$</th>
<th>$B_{stable}$</th>
<th>$B_{mono}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>35.9464</td>
<td>5.5034</td>
<td>35.2931</td>
<td>28.8247</td>
</tr>
<tr>
<td>$C_2$</td>
<td>25.1069</td>
<td>21.4572</td>
<td>86.9245</td>
<td>83.1662</td>
</tr>
<tr>
<td>$C_3$</td>
<td>385.169</td>
<td>290.014</td>
<td>1622.99</td>
<td>1548.22</td>
</tr>
</tbody>
</table>

Table I: BUDGET LIMITS FOR DIFFERENT COST STRUCTURES

bears some of the characteristics of both parents (crossover) along with certain new characteristics (mutation). The crossover part ensures that we perform a local search around both parents, whereas the mutation part ensures that we are not stuck in a local optima. Since we have generated the initial population based on differential smudges, two individuals can be easily combined as long as their smudges do not overlap; thus implementing the crossover operation. For overlapping smudges, we perform a randomized mixing to take care of mutation. This indicates another advantage of using the smudge based population generation, wherein the process of combination of parents has a physical meaning: the combination of smudges.

ii) Quality of children: The fundamental complication of the problem dealt with in this work is the multi-objective optimization of interlink structures, wherein we are interested in obtaining the $\rho$ which leads to the maximum $\psi$ while satisfying the budget constraint. Since we generate the initial population starting from the $\rho$ corresponding to $B_{int}$, our population has an implicit bias towards interlink structures of high $\psi$. Thus during the inspection of quality of the population, we initially focus on removing the $\rho$’s with costs higher than $B_{int}$ from the population. Once we have reached the stage where a majority (chosen to be 85%) of population satisfies the budget constraint, we again focus on removing interlinks with low $\psi$ from the population.

iii) Convergence of population: The convergence of genetic algorithms is largely an open problem with various works using various heuristics to overcome this issue. The trade-off in obtaining a decision with regard to convergence is between the quality of the final solution and running time of the algorithm. For our case, the population is assumed to have converged when an entire generation of children are produced without any significant increase in the quality of the population.

VI. EFFECT OF COST MATRIX STRUCTURE

In the previous two sections, we have discussed and designed various algorithms for obtaining interlink structures satisfying a budget constraint. It must however be noted that the actual structure of the cost matrix ($C$) would play a very important role in determining the performance of the algorithms. A mathematical analysis of the effect of various structures of $C$ would undoubtedly be a very interesting future work, which we are planning to pursue as well. However in this section, we will present the performance of the algorithms for three cost matrices:

C1: Uniformly generated cost matrix, where the cost matrix elements $[c_{ij}]$ are chosen randomly between 0 and 1.

C2: Noisy rows, where the first row of $C$ is generated uniformly randomly and remaining rows of the matrix are noisy versions of the first row.

C3: Uniform cost matrix with different row means, where the cost matrix rows are generated as in C1 with the difference that the row means are chosen randomly between 0 and 100.

Next, we discuss the physical interpretation of these cases. C1 denotes the simplistic case where all rows have the same statistical properties (uniform distribution) but are independent of each other. C2 represents the case of correlation between the various rows, i.e. if the cost of interlinking a particular node $a \in A$ to $b \in B$ is high, the cost of interlinking all layer $A$ nodes to $b$ will most likely be high. This captures the property of high or low inherent cost, where nodes of a layer have a tendency to retain their costs when interlinked to different nodes of the other layer. C3 denotes a generalization of C1, wherein the row means of $C$ are different, representing the case where certain interlinking costs can be much higher or lower than others. It can
be clearly observed from Table I that the cost of a blind (random) monotonic arrangement ($B_{\text{mono}}$) is much higher than both $B_l$ and $B_u$. The values in the table are averaged over 100 runs and the corresponding box plot is presented in Fig. 2. Table I portrays the key conclusion of this work– designing optimal interlink structures while being agnostic about the cost of interlink construction leads to a very high network construction cost. Furthermore, the difference between $B_u$ and $B_{\text{mono}}$ indicates that the maximum robustness of the network can be achieved at a much lower costs ($B_u$), if the cost structure is considered during network design. With respect to the cost matrices $C_1$, $C_2$, and $C_3$, it can be observed from the table that for all these cases the SMA leads to a much tighter approximation of $B_l$ as compared to the naive algorithm. Furthermore, when the costs for interlinking different nodes are uncorrelated ($C_1$) there is less difference between the naive and SMA algorithm (8%), while for correlated cases ($C_2$) the difference is higher (17%). This indicates that naive algorithm gives a tighter approximation for $B_l$ under uncorrelated cost structure than a correlated cost structures. As a part of future work, we plan to try deriving more general relationships between the cost structure and performance of the approximation algorithms.

Next we present the performance of the algorithms designed for obtaining optimal interlink structures for intermediate budgets. We have provided results for the uniform cost matrix structure represented by $C_1$; the results for the other cases are qualitatively similar. Note that we are focusing on a particular network configuration (initial attack on 20% nodes) where interlink structures for intermediate budgets lead to $\psi$ in the range of (0.66, 0.75). A general trend that can be observed from Fig. 3 is that the robustness of interlink structures for intermediate budgets approaches the maximum robustness at a rate much higher than linear. This further indicates the usefulness of considering cost structure for the construction of interlinks– even for intermediate budgets between $B_l$ and $B_u$, we can design interlink structures which can achieve close to optimal robustness. In terms of performance comparison of the three algorithms, it can be observed that the greedy cost-effective is the best, while the genetic and degree-based algorithms are better than each other in different regimes. This also indicates a very interesting result that genetic algorithms are not better than the proposed greedy algorithms even though they are computationally more expensive. It is important to clarify at this point that there are numerous variants of the genetic algorithm and we have only worked with a small subset of those. Thus we consider the inability of genetic algorithms to produce good results as a preliminary finding, which we plan to further investigate as a part of future work. However an intuition behind this unexpected result can be developed. The crossover and mutation operations lie at the heart of genetic algorithms and perform the task of
local and global searching, respectively. Genetic algorithms work well when small perturbation in parent genes lead to small changes in the performance of child genes, thus achieving the local search criteria. However for our case, after modification of these operations as discussed in Section V-C, the resulting children significantly differ from the parents in performance due to the coupled factors of interlink construction costs and interlink arrangements. Thus the genetic algorithm behavior is similar to that of a random search providing a possible explanation for the poor performance.

VII. Simulation Results

We have designed a Python based simulation framework, to examine the actual network robustness for intermediate budgets \( B_{int} \in [B_l, B_u] \), where the algorithms presented above are used to construct the interlink structures. The simulation framework utilizes the NetworkX library to construct the constituent layers as Erdos-Renyi (ER) graphs of size \( N = 500 \) and thereafter emulates the cascading failure process by obtaining the mutually connected component at each iteration until a steady state is reached. The results are qualitatively similar for other graph structures like scale free graphs as well, and ER graphs were chosen for their popularity in network science literature [3, 6, 7]. Note that the maximum and minimum robustness correspond to the monotonic and anti-monotonic arrangement of nodes, respectively, on the basis of intra-layer degrees [5, 7]. We have taken the budget constraint to be \( B_{int} = B_l + \frac{B_u-B_l}{5} \) to depict a low budget condition and it can be observed from Fig. 4 that the robustness of the cost-effective algorithm is quite close to the optimal robustness, whereas the degree based algorithm does not perform very well. This corroborates the result in Fig. 3, wherein we can observe that for low budget conditions degree-based and cost-effective algorithms are worst and best, respectively.

VIII. Conclusion and Future Work

In this paper, we have considered the effects of cost constraints on the design of optimal interlink structures to maximize the robustness of interdependent networks against node failures. In adherence to our initial claim, this work demonstrates that interlink structures can be designed at a much lower cost \( B_u \) whilst maintaining the optimal performance of the resulting network. Furthermore, even for intermediate budgets between \( B_l \) and \( B_u \), we have shown, through our algorithms and simulation results, that we can design interlink structures with close to optimal robustness. Our main focus in terms of future work is to generalize the network model from various perspectives, to solve the problem for a wide variety of interdependent networks. A mathematical framework capturing the effect of cost structure on network design represents another avenue of future research in this area. Although this work represents the initial footsteps towards developing a rigorous understanding of the impact of cost constraints on network design, we believe that the algorithms and analysis herein establish the importance of cost constraints on network design and would lead to further development of optimal interlink structure design under budget constraints.

REFERENCES


