

# The impact of mobility models on the performance of P2P content discovery protocols over mobile ad hoc networks

Hamideh Babaei · Mahmood Fathy · Reza Berangi · Morteza Romoozi

Received: 31 January 2012 / Accepted: 6 November 2012 / Published online: 20 November 2012  
© Springer Science+Business Media New York 2012

**Abstract** One of the challenges over mobile ad-hoc networks is content discovery. P2P content discovery techniques including structured and unstructured can be employed in MANETs by considering its special characteristics and limitations. The most important characteristic of MANETs is the mobility of the nodes which creates a dynamic topology. A novel framework is presented to evaluate the effect of mobility and its models on the resource discovery. By using several metrics, this framework is capable of evaluating the effect of mobility on the underlay structure and subsequent changes on the overlay structure. The results obtained from extensive simulation, presented here, clarify the significant role of the mobility models on the performance of P2P content discovery protocols. These results are supported by mathematical analysis of content discovery protocols.

**Keywords** Mobile Ad-hoc Networks · P2P networks · Mobility models · Analysis framework · P2P content discovery protocols · Mathematical analysis · Generalize random graph

---

H. Babaei (✉) · M. Romoozi  
Department of Computer Engineering, Science and Research  
Branch, Islamic Azad University, Tehran, Iran  
e-mail: h.babaei@srbiau.ac.ir

M. Romoozi  
e-mail: m.romoozi@srbiau.ac.ir

M. Fathy · R. Berangi  
Department of Computer Engineering,  
Iran University of Science and Technology, Tehran, Iran

M. Fathy  
e-mail: mahfathy@iust.ac.ir

R. Berangi  
e-mail: rberangi@iust.ac.ir

## 1 Introduction

Mobile ad hoc networks (MANETs) are wireless networks with no fixed structure [1], which have been presented as a flexible, simple, scalable technique for creating connection networks, collecting information and distributed processing of information. Rescue operations and battle fields are among important scenarios that are presented for MANETs. The most critical limitation of such scenarios is the lack of a stable and predefined structure. This limitation obstructs applying client–server architectures in network based applications and the network becomes vulnerable for peer to peer applications so that each node plays the role of both client and server simultaneously.

Another significant property of MANET is the mobility of participating nodes which causes changes in the network topology. The manner of node movement can be described in computer simulation by network mobility model.

On the other hand, today peer to peer (P2P) networks have also had considerable growth compared with client–server architectures and have dedicated high traffic to themselves [2]. The most important applications of this network are sharing information and resources, content dissemination, peer-to-peer connections, collaborative programs and content discovery [3]. In these networks two nodes can exchange information without any need to an infrastructure or a central node [3].

P2P and MANETs have many similarities in nature, including being self-organized and decentralized and having dynamic topology. However, there are differences between them too [4–7]. For instance, topology in MANETs is more dynamic than that in P2P networks due to node movement in MANET. In addition, MANETs encounter limitations such as low computing capacity, radio transmission range and consumed power.

The similarities between P2P networks and MANETs, the significance and simplicity of MANET deployment and wide applications of P2P have opened a new research field that attempts to apply overlay structure of P2P networks in the application layer of MANETs. However, before applying P2P networks over MANETs, it is critical to evaluate the effect of the differences between MANET and P2P networks on the performance of P2P applications.

Accordingly, this article aims to evaluate the movement characteristics of MANET as mobility models on the efficiency of P2P applications. To achieve this aim, it presents a framework by which not only can it answer the question of whether mobility influences P2P applications, but also it can evaluate the impact of different mobility models and the way the effects are created.

Since there are a variety of mobility models and various applications for P2P networks, the framework inputs have been selected in a way that can be generalized to other models and applications. Therefore, for evaluating this framework, Random Waypoint mobility model (RW)[8], Obstacle Mobility model (OM)[9], Manhattan mobility model [10] and Reference Point Group Mobility model (RPGM)[11] have been selected as the inputs of mobility models and Chord [12], Random walk [13] and Gnutella [14] have been selected as content discovery protocols. They have been selected in a way that each of selected mobility models is the symbol of a category of mobility models, including realistic, random, entity or group mobility model [15]. Input content discovery protocols have been selected from main categories of protocols including unstructured and structured ones.

The proposed framework is comprised of three main phases: 1. Mobility models and its quantitative and qualitative evaluation metrics; 2. Underlay structure and its evaluation metrics (underlay structure here refers to network protocol stack except application layer and its protocols); 3. Overlay structure and its evaluation metrics (overlay structure here means application layer that is equipped with P2P connection protocols).

After presenting the framework in details, the article validates all the relevant metrics and analyzes their results. The analysis clarifies the relationship between mobility and underlay structure. Consequently, the relationship between underlay structure and overlay structure is clarified too. Finally, by using the results of the analyses and based on the transition rule, the relationship between mobility and overlay structure is inferred.

Eventually, for proving revealed relation among under-investigation-phases in analysis framework, mathematical model is proposed for modeling content discovery protocols. Based on this model, the claimed relations in simulation results are proved. In the modeling, overlay network is considered as a random graph (GRG)[16]. So GRG has been used for modeling.

This article is organized as follow: first, previous studies on the evaluation of P2P networks over wireless networks and on the effect of mobility on the performance of MANETs are reviewed. Then, it presents mobility models and their classifications. Next, P2P applications and their various architectures are explained. The proposed evaluation framework is discussed in details and is evaluated for four sample mobility models under different overlay structures. Finally, a mathematical model is proposed for modeling content discovery protocols and the results of the previous section are proved.

## 2 Related works

Previous research which is related to this article can be divided into two branches. First, the research which evaluated and improved the performance of MANET based on its mobility pattern. Second, the studies which evaluate the performance of content discovery techniques over MANETs.

### 2.1 Evaluation of the performance of MANET protocols under different mobility models

Studying the effect of mobility models, as a key component in MANET on routing protocols, is another appealing research area. For instance, authors have presented an analysis framework, in which the impacts of RPGM, Manhattan, Freeway [10] and Random Waypoint mobility models on DSR [17], DSDV[18] and AODV[19] routing protocols are evaluated based on throughput and routing overhead metrics [10].

DSR and AODV protocols as reactive protocols and DSDV protocols as proactive ones, under the Random Waypoint mobility model and based on the metrics of control overhead and Packet delivery ratio have been analyzed in [20]. The authors concluded that reactive protocols present better performance than other protocols in high mobility rates. However, the superiority of DSDV is completely evident in rather static networks. Similarly, new mobility models have also been presented that are distinguished from other mobility models by introducing a new metric; then popular routing protocols of MANET have been evaluated over them [21].

### 2.2 Evaluation of the performance of P2P content discovery techniques over MANET

Similarities and differences of MANET and P2P networks have been discussed in [22] and [23]. The studies show that the protocols that are developed for wired networks have poor efficiency for MANET because equipment in MANET are prone to errors and its topology suffers from high level of dynamicity [24]. Accordingly, it seems necessary to develop or adapt P2P protocols to make them suitable for the characteristics of MANETs.

In [6] Dynamic P2P Source Routing (DPSR) has been proposed. It is a routing ad hoc protocol that uses the strategies of DSR routing protocol and P2P Pastry protocol.

In [5], authors have concentrated on minimizing the highly dynamic effect of MANET topology for P2P networks. They have also presented algorithms for reconfiguration of P2P topology. In [25], MPP protocol has been proposed for coordinating virtual and physical topologies so that it can interlink its own routing layer with the underlying physical layer.

Bit Torrent-like protocol for applications over MANET has been proposed in [26]. Attempts have also been made to create a virtual collaborative environment over MANET for military and emergency applications [27].

As mentioned earlier, a lot of studies have been conducted on P2P protocols and their applications in MANET, but there are a few studies on the effects of MANET characteristics, such as node mobility, link quality etc., on the performance of P2P applications. However, having such knowledge is helpful in extracting more intelligent protocols and in identifying key factors that affect the performance of a certain scenario or deployment. Such studies have been done on a theoretical level [7].

However, authors implemented and evaluated content discovery P2P protocols, Gnutella, Random walk and Chord over MANET under the mobility model of Random Waypoint and based on energy consumption, response time and hit rate metrics [28]. They have come to the conclusion that since unstructured P2P protocols send messages by employing flooding method, they show a better hit rate with a higher cost in a highly dynamic MANET with low reliability. In contrast, structured protocols that send a query message for each request do not yield efficient results due to the frequent loss of messages in such environments. The authors presented algorithm for improving these protocols [29]. The improvement is done in structured protocols by increasing the number of queries and in unstructured protocols by increasing the probability of choosing neighbors with fewer loads for forwarding messages.

### 3 Mobility models in MANET

Mobility models are classified from different aspects. In a popular classification, they are divided into two categories of *Entity* and *Group* [15], based on the amount of movement interdependency between mobile nodes. In another classification, mobility models are classified into two categories of *Random* and *Realistic* [30], based on the similarity of node movement to its movement in real environment.

#### 3.1 Entity mobility model

These models describe how a node moves individually without any attention to other nodes around it. That is, a

node's choice of destination, speed and route is not influenced by the movement of other nodes. Random waypoint is an example of these models where a mobile node randomly chooses a destination and a speed from  $[V_{\min}, V_{\max}]$  and moves toward it. In destination, it randomly pauses from  $[P_{\min}, P_{\max}]$  and repeats the procedure again.

#### 3.2 Group mobility model

In these models, nodes' behaviors are dependent on each other and the nodes do not make decisions as an independent entity; rather they follow group movement. An example of these models is Reference Point Group Mobility Model (RPGM). In RPGM model, the nodes are considered as groups in each of which there is a reference point which can be a geographic, symbolic point or a leader node. This point controls movement behavior of the group.

#### 3.3 Random mobility models

This group of models considers node movement completely randomly. Neither environmental factors, such as buildings, nor non-environmental factors like movement rules can limit the node mobility. Random waypoint, Random walk and Random direction mobility model belongs to this category [15].

#### 3.4 Realistic mobility models

Unlike random models, in realistic models some limitations are imposed on node movement. The limitations may be due to environmental obstacles, such as buildings, or to the rules made for node movement such as moving in predefined pathways. The rules are made to make node movement more similar to real nodes. Realistic mobility models usually contain the three following sub-models: environment sub-model, signal obstruction sub-model and movement pattern sub-model [30].

Obstacle Mobility model [9] is one of these models. In this model, first obstacles are defined. Then pathways are formed by Voronoi graph [31]. Each node randomly selects a vertex of graph as its destination. Then it goes there through the shortest path to it. In signal obstruction sub-model of this model, each node considers an inhibition cone for each obstacle around itself. If a node is present in an inhibition cone, cannot communicate with that node.

Another type of realistic mobility models is concerned with describing automobile movements in streets. These models can be applied in Vehicular Ad-hoc Networks (VANET) [32]. An instance of these models is Manhattan mobility model. In the Manhattan mobility model, first a grid of streets is defined for it so that mobile nodes can move only in these streets. The model employs a probabilistic approach in the selection of

nodes' movements because when they arrive at the intersection, they change their directions with 25 % probability and continue moving with 50 % probability.

#### 4 Content discovery protocols

One classification of content discovery protocols [3] divides them into two categories of structured and unstructured categories. In the following parts, each of the categories is discussed in details.

##### 4.1 Unstructured protocols

In these protocols, there is no specific rule to define the location of stored data and network topology. They use query flooding to search. Gnutella and Random Walk are two significant protocols of this type.

Gnutella protocol that has been presented in [14] in details, propagates queries by controlled flooding. When a node receives a query, it examines whether it has the requested content or not. If it does, the node sends back a result message to the query source; otherwise, the node sends the query to all the nodes in its neighborhood. To prevent the infinite distribution of the message, the value of a TTL field that exists in the query message is reduced per hop. The messages which have zero TTL or duplicated messages are discarded.

In Random walk protocol [13], the node that sends a query creates  $n$  walkers or the same query messages with fixed TTL. Then, it selects  $n$  neighbors from among its neighbors with a uniform probability and sends a walker to each. When a peer receives a walker, it examines whether it itself has information or not. If it does, the walker is deleted from the network and the peer starts creating connection with the query source. But if the walker itself does not have information and its TTL is not zero, it forwards the walker to a random neighbor and reduces one of its TTL. In this protocol, one cache is used to avoid accepting one query several times.

##### 4.2 Structured protocols

In structured protocols, the neighborhood connection between nodes and the location of stored data are clearly defined. Chord is one of the most well-known structured protocols has been presented in [12] in details. It uses hash functions to divide files and content among peers. In this protocol, all the peers are arranged based on their IP addresses and create a ring. Each peer is responsible for files having hashes within the range  $[\text{hash}(\text{my-ip}), \text{hash}(\text{my-successor-ip})]$ . This protocol performs the lookup by using *finger table*. Every  $i$  entry in finger table shows a peer that contains files whose highest value of hash is  $\text{hash}(IP_i)$  where  $\text{hash}(IP_i) \leq (\text{myhash} + 2^i) \bmod 2^m$ ,  $m$  is the number

of bites generated by hash function and *myhash* is the hash of the IP of current peer. This relation guarantees that every query will be solved at approximately  $\log(n)$  forwarded message by using a forward algorithm that implements a binary search in the hash space.

#### 5 Analysis framework

This article presents a framework for analyzing effect of mobility on the performance of P2P content discovery protocols which contains three interdependent phases. In the first phase, mobility models and the related metrics are examined. In the second phase, the underlay structure which contains network protocol stack (MAC and routing layer) is evaluated. In the third phase, the overlay structure containing P2P content discovery protocols are investigated. There is a middle phase between every two subsequent phases whose main task is to analyze the relation between them. In this way, the relation between mobility models and underlay structure on the one hand and the relation between underlay structure and overlay structure on the other hand are analyzed. Then, according to transition rule, the relation between mobility and overlay structure is found out. In the following sections, the components of the framework are discussed in details, as illustrated in Fig. 1.

##### 5.1 Mobility phase

In this part of the framework, mobility models and their major characteristics are identified. Hence, four mobility models enter this section as input. As it was stated previously, these four models are chosen in a way that each represents a class of mobility models. The models are Random Waypoint, Reference Point Group Mobility, Manhattan and Obstacle Mobility Model.

##### 5.2 Mobility metrics

###### 5.2.1 Geographical uniform distribution

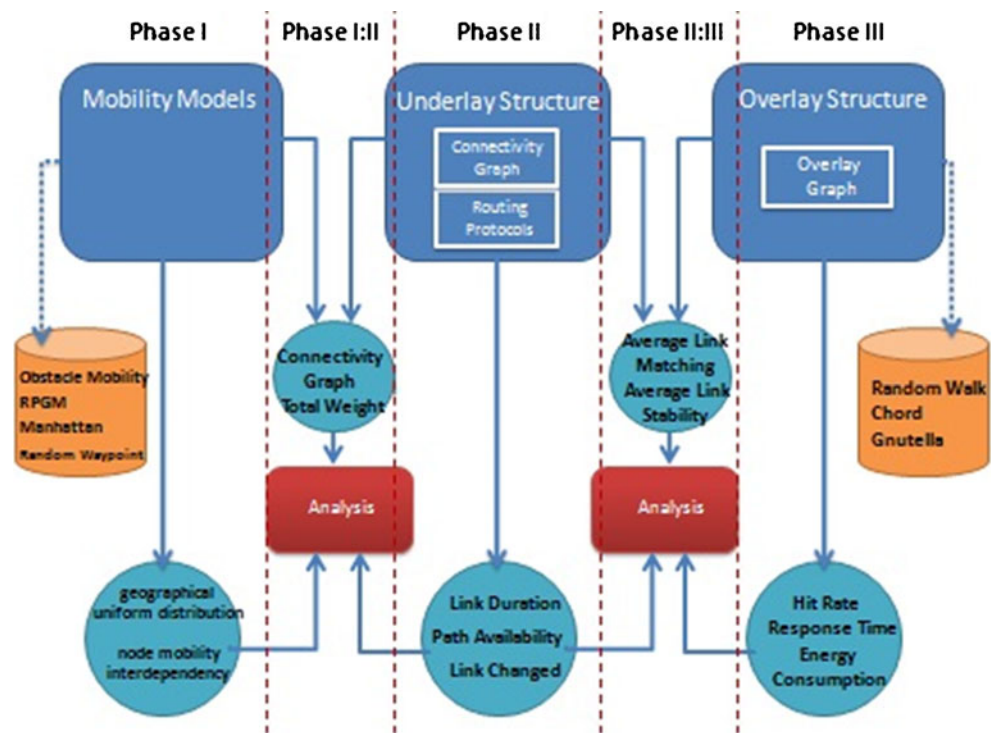
This metric determines how uniformly the nodes have distributed in simulation terrain. The higher the metric is, the more uniformly the node density is distributed. In contrast, less value of the metric shows that nodes are gathered in smaller space of the terrain.

To calculate it, simulation environment is considered as a  $k \times k$  grid and the Expected value that each node might be located in each of the grid cells is estimated.

$$E(i, t) = \sum_{j=1}^k \sum_{l=1}^k \sqrt{(x_i - x_j)^2 + (y_i - y_l)^2} * p(x_j) * p(y_l) \quad (1)$$



Fig. 1 The analysis framework



$$p(x_j) = \sum_{t=1}^T f / T \cdot f = \begin{cases} 1 & \text{if } f \text{ node exist in } x_j \text{ scope} \\ 0 & \text{if } f \text{ node not exist in } x_j \text{ scope} \end{cases}$$

In Eq. 1,  $E(i, t)$  is the Expected value of the node  $i$  distribution in the grid space. It is obtained from multiplying the node  $i$  distance from each grid cell  $(j, l)$  by the probability of node  $i$  existence in that cell.  $P(x_i)$  is the probability of node existence in horizontal zone  $x_j$  and  $P(y_i)$  is the probability of node existence in vertical zone  $y_l$ .

This probability is calculated by post processing of simulation snapshots that are captured every  $t_s$  seconds. These snapshots identify node location in a discrete event simulation periodically.

$$AGD = \frac{\sum_{i=0}^N \sum_{t=1}^T E(i, t)}{N * T} \tag{2}$$

Finally, Average Geographic Distribution (AGD) can be calculated by summing  $E(i, t)$  for each  $n_i, i \in [0, N]$  ( $N$  is the number of nodes) and each snapshot  $t, t \in [0, T]$  ( $T$  is the number of snapshots) and dividing it into  $N * T$  (Eq. 2).

### 5.2.2 Node movement interdependency

By using this metric, node movement interdependency can be described, i.e. the impact of a node movement on other nodes and vice versa. To calculate this metric, first  $d_{i/t-1}(i)$  is defined (Eq. 3). This variable shows the movement of node  $i$

at time  $t$  as compared to its own previous location at time  $t-1$ . Each  $t$  time denotes a snapshot of simulation that captured nodes' location periodically.

$$\forall t \in T, \forall i \in N \quad d_{i/t-1}(i) = \sqrt{(x_i(t) - x_i(t-1))^2 + (y_i(t) - y_i(t-1))^2} \tag{3}$$

Where  $x_i(t), y_i(t)$  is the coordination of node  $i$  in snapshot  $t$ . The result of calculating  $d_{i/t-1}(i)$  for all nodes would be a  $T \times N$  matrix, where  $N$  is the number of nodes and  $T$  is the number of simulation snapshots. Every column of this matrix would belong to one node. As the next step, we calculate correlation coefficient of the movement of each node in relation to other nodes by using Multiple Correlation Coefficient Pearson Regression [33]. This coefficient can be described as the dependency of that node's movement on other nodes.

Finally, by calculating the average of these values for all nodes, the Average interdependence of nodes' movements is estimated.

### 5.3 Underlay structure phase

Underlay structure refers to the network underlay on which overlay structures of p2p networks are implemented. It refers to network protocol stack including MAC, Routing and transmission layer. But in this framework, we consider

underlay structure as a connectivity graph (CG) that is defined as follow:

$$CG = (V, E) \text{ where } |V| = N \text{ and } link(i, j) \in E \text{ if } D_{ij}(t) \leq R \quad (4)$$

$N$  is the number of wireless nodes; Between the two nodes there is a link or edge if  $D_{ij}(t)$  (Euclidean distance between them) is less than  $R$  (transmission range of the nodes). In this case, the two nodes are neighbors.

By considering underlay structure as a graph, MAC and routing can be described simultaneously. It is because the neighborhood condition in Eq. 4 denotes MAC neighboring

method. In addition, finding the shortest path in the graph can be considered as a type of routing protocol.

#### 5.4 Underlay structure metrics

##### 5.4.1 The average number of link changes

To calculate this metric, first  $LC(i, j)$  which is defined as the average number of times that the link between the two nodes of  $i$  and  $j$  is changed (connected or broken), is obtained. In the following equation (Eq. 5)  $LC_{t/t-1}(i, j)$  is equal to 1 if the link between the node  $i$  and  $j$  change from connected to broken or vice versa in snapshot  $t$  comparing to snapshot  $t-1$ .

$$LC(i, j) = \sum_{t=0}^T LC_{t/t-1}(i, j) \begin{cases} LC_{t/t-1}(i, j) = 1 & \text{if } link(i, j)_t \in E \text{ and } link(i, j)_{t-1} \notin E \\ & \text{or } link(i, j)_t \notin E \text{ and } link(i, j)_{t-1} \in E \\ LC_{t/t-1}(i, j) = 0 & \text{o.w} \end{cases} \quad (5)$$

Finally Average number of link change (ALC) can be calculated according to the Eq. 6.

$$ALC = \frac{\sum_{\forall(i, j) \in V} LC(i, j)}{N(N-1)/2} \quad (6)$$

##### 5.4.2 Average link duration

This metric refers to the average time that links remain connected during simulation. To calculate this, first a fraction of time that each link has been connected during simulation is measured, and then the average is estimated for all the possible links.

In the Eq. 7,  $LD(i, j)$  is the sum of the number of snapshots that a link was established between node  $i$  and  $j$ . if an edge (link) exist between  $i$  and  $j$  then  $la(i, j) = 1$ . ALD is the average value of link duration for all the possible links (Eq. 8).

$$LD(i, j) = \sum_{t=1}^T la(i, j, t) \begin{cases} la(i, j, t) = 1 & \text{if } e_{(i, j)_t} \in E \\ la(i, j, t) = 0 & \text{o.w} \end{cases} \quad (7)$$

$$ALD = \frac{\sum_{\forall(i, j) \in V} LD(i, j)}{N(N-1)/2} \quad (8)$$

##### 5.4.3 Average path availability

This metric refers to the average time during which there is a path between the two nodes of  $i$  and  $j$ . To calculate it, a fraction of simulation time during which there is a path between the two nodes is considered (Eq. 9). If there is a path between nodes  $i$  and  $j$   $pa(i, j, t) = 1$ . Then, the average of this value for each node is estimated (Eq. 10).

$$PA(i, j) = \sum_{t=1}^T pa(i, j, t) \begin{cases} pa(i, j, t) = 1 & \text{if there is a path between } i \text{ and } j \\ pa(i, j, t) = 0 & \text{o.w} \end{cases} \quad (9)$$

$$APA = \frac{\sum_{\forall(i, j) \in V} PA(i, j)}{N(N-1)/2} \quad (10)$$

In Eq. 9  $pa(i, j, t)$  equals 1 if there is at least one potential path between the nodes  $i$  and  $j$  in snapshot  $t$ .  $PA(i, j)$  sums number of snapshots that a path between node  $i$  and  $j$  can be found. Average Path Availability (APA) can be calculated by averaging the values of  $PA(i, j)$  for all nodes.

#### 5.5 Connectivity graph average total weight

This section of framework is a part of the middle phase of mobility models and underlay structure. This metric clarifies the relation between mobility and connectivity graph. To calculate it, the structure of Connectivity graph at each snapshot is taken into account, and the weight of the connection link between the nodes of  $i$  and  $j$  is measured based on Eq. 11.

$$w_{CG}(i,j) = \frac{\sum_{t=0}^T \text{dist}(i,j,t)}{T} \quad (11)$$

$w_{CG}(i,j)$  is the average distance between the two nodes of  $i$  and  $j$  during simulation.  $\text{dist}(i,j,t)$  is Euclidean distance between nodes  $j$  and  $i$  in snapshot  $t$ . The total weight of connectivity graph equals the total weight of all the edges of the graph. Finally, connectivity graph average total weight (CGTW) is calculated by averaging total weight of all links of connectivity graph for all snapshots (Eq. 12).

$$CGTW = \frac{\sum_{t=1}^T \sum_{\forall (i,j) \in E} W_{CG}(i,j)}{T} \quad (12)$$

The less the value of this metric, the higher the performance of connectivity graph will be. Because the less the average distance between the two vertices of edge in the graph during simulation, the less the possibility of breaking this edge will be. The graph will consequently be more stable too.

## 5.6 Overlay structure phase

Overlay structure can be considered as a graph which is known as overlay graph. Vertices of this graph are peers participating in application, and its edges represent virtual neighborhood in the overlay structure. This overlay graph (OG) can be defined as follow:

$$OG = (V, E) \text{ where } |V| = P \text{ and } \text{link}(i, j) \in E \text{ if } i \text{ (13)}$$

*is neighbor of j or vice versa*

$P$  is the number of peers.

## 5.7 Overlay structure metrics

Overlay structure metrics are determined based on the type of P2P application. Thus, in this phase, we define and calculate the most important content discovery protocol metrics as follow:

1. Hit Rate: the number of queries that are responded successfully in P2P networks
2. Response Times: the value of time that a peer should keep waiting for receiving the required its query response. The time contains the value of time needed for query to be transmitted in the network, to discover the target content and to return the response back to the peer.
3. Energy Consumption: the average energy that each node consumes during receiving and sending signals.

## 5.8 Overlay graph total weight

This part of framework is the middle phase between the two phases of underlay and overlay structures. It contains

metrics whose values are calculated by considering the target graphs in both phases, i.e., connectivity graph and overlay graph. The metrics' values determine the effect of changes in connectivity graph on the overlay graph.

Each node in overlay graph is equivalent to a node in connectivity graph. However, each edge in an overlay graph can correspond to one or several edges (i.e. a path) in the connectivity graph.

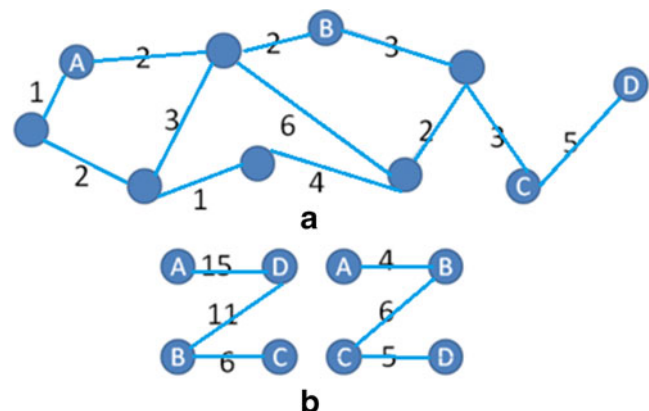
But the point here is that the lack of matching between the two graphs influences the performance of P2P applications, and decreases P2P network throughput [34]. Figure 2 illustrates this matter clearly. In the figure 2(b), there are two layouts for constructing an overlay graph by nodes A, B, C and D. But their total weight which is calculated based on connectivity graph is different. As the weight of each link in connectivity graph is communication cost, the overlay with minimum total weight is desired.

Accordingly, in this part of the framework we present two metrics that aim at calculating the amount of matching between these two graphs in terms of Average Link Matching and Average Link Stability.

### 5.8.1 Average link matching

This metric clarifies how a link in an overlay graph matches a corresponding path in connectivity graph. For example, if a link is established between nodes  $i$  and  $j$  in overlay graph, there is a path between  $i$  and  $j$  in connectivity graph. The path length in the best case would be 1, which means the link between  $i$  and  $j$  in an overlay graph is exactly match connectivity graph. Therefore, the less the path length is, the more the link in an overlay match connectivity graph.

To calculate this metric, first the weight of each link in an overlay graph is calculated according to Eq. 14 where  $d_{\text{shortest\_path\_CG}}(i,j,t)$  is the length of the shortest path between nodes  $i$  and  $j$  in connectivity graph in



**Fig. 2** a Connectivity graph (underlay structure). b two possible Overlay graphs of connectivity graph (a)

snapshot  $t$ .  $c$  is a constant which represents cost of each link in the connectivity graph based on delay, traffic, etc.

$$w_{OG}(i,j) = \frac{\sum_{t=0}^T d_{shortest\_path\_CG}(i,j,t) * c}{T} \quad (14)$$

Finally, Average Link Matching (ALM) can be calculated by Eq. 15. It first computes the total weight of the overlay graph based on the calculated weight of each link in Eq. 14, then divides it into the number of link in overlay graph.

$$ALM = \frac{\sum_{\forall i,j \in V} w_{OG}(i,j)}{|E|} \quad (15)$$

### 5.8.2 Average link stability

This metric can represent the stability of overlay graph links against the changes in underlay structure. As mentioned earlier, underlay structure, as a connectivity graph, changes continuously because of topology changes resulting from nodes' movements; its neighbor nodes are frequently changing too. This can lead to an increase in the cost of a virtual link between two nodes in the overlay graph, which is based on the distance between nodes in connectivity graph. It can also cause the disconnection of a link in the overlay graph. The reconfiguration of overlay structure, which is a costly task results [34]. Therefore, it necessitates presenting a quantitative metric that estimates the stability of a virtual link in overlay structure. On the other hand, this metric can demonstrate the relation between overlay and underlay structures.

According to Eq. 16, for calculating this metric, the weight of each link of the overlay graph is calculated based on availability of that link in the underlay graph in each snapshot.  $w_{OG}(i,j)$  is the average weight between  $i$  and  $j$  in total simulation time and Average Link Stability (ALSta) obtains the average  $w_{OG}(i,j)$  for all overlay graph edges (Eq. 17).

$$w_{OG}(i,j) = \frac{\sum_{t=0}^T PA(i,j,t)}{T} \quad (16)$$

$$ALSta = \frac{\sum_{\forall i,j} w_{OG}(i,j)}{|E|} \quad (17)$$

## 6 The analysis and validation of the framework

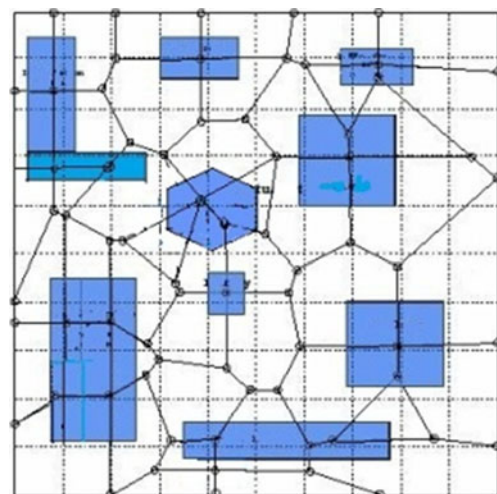
In this section, the proposed framework whose inputs are four mobility models and three p2p content discovery protocols is evaluated and its validity is examined. The effect of mobility on underlay structure and consequently on overlay

structure is also studied both quantitatively and qualitatively. To this end following steps should be considered.

1. Validating the mobility metrics and evaluating the differences of mobility models based on these metrics
2. Validating the metrics of underlay structure, evaluating and analyzing the relation between these metrics.
3. Validating the metric of the average weight of connectivity graph, and evaluating the differences in its value under different mobility models.
4. Evaluating and analyzing the relation between mobility and underlay structure based on the production diagrams in sections 1, 2, and 3.
5. Evaluating and validating overlay structure metrics and their relation.
6. Validating the metrics of Average Link Matching and Average Link Stability and analyzing the results based on different mobility models.
7. Evaluating and analyzing the relation between underlay and overlay structures based on diagrams obtained from sections 2, 5, and 6.

### 6.1 Network characterization

To simulate, one of the most common network simulators, i.e. NS-2 is used. This simulator can simulate MANETs under different scenarios. In order to generate scenarios extracted from mobility models of Random Waypoint, RPGM and Manhattan, Bonnmotion [35] which is a Java Base tool has been used. The scenario of Obstacle Mobility model has also been generated by using the tool that the authors have presented [9]. Obstacle Mobility model also uses a terrain including obstacles and pathways illustrated in Fig. 3. Given that, none of the other mobility models



**Fig. 3** Obstacles and the Voronoi graph created using the corners of the obstacles



support signal obstruction mechanism, this component has been deactivated for Obstacle Mobility model. Each scenario has been run 20 times for each protocol. Characteristics of mobility models, underlay structure and overlay structure is presented in Table 1.

In the underlay structure nodes use IEEE 802.11b, one of the most common MAC protocols in wireless networks, upon Cisco AiroNet350 [36] with transmission power of 10 dBm. Its consumption energy is 1.6887 W for transmission, 0.6699 W for idle, and 1.0791 W for reception modes.

As it was mentioned earlier, three protocols of Chord, Gnutella and Random walk have been simulated. In Chord, finger tables are updated every 5 s and *stabilize* functions are run every 10s and *ping* messages are sent every 10s. In Gnutella it is assumed that each node has a maximum of six neighbors. TTL of messages is set to 4 and ping messages are sent every 10 s. For Random walk the number of neighbors is assumed to be 6 and every query runs four walkers, TTL of each walker equals 1/4 of the number of nodes. The packet length is considered 64 bytes.

## 6.2 Validating and evaluating mobility metrics

### 6.2.1 Geographical uniform distribution

Figure 4 illustrates the degree of uniformity in geographic distribution of nodes in the environment for different mobility models with various speed and various number of nodes respectively. Unrealistic mobility models which are not dependant on predefined pathways, such as RPGM and RW distribute nodes in the environment uniformly. It should be mentioned that, compared with RW entity mobility model, RPGM model with group mobility has lower results.

Manhattan and OM realistic mobility models have similar geographic distribution; OM model has better results

compared to Manhattan mobility model. Because when arriving to a cross, nodes in Manhattan mobility model are inclined to have direct movement. So nodes reach the environment boundary rapidly. Therefore based on Eq. (2), Geographical uniform distribution metric value for Manhattan mobility model increases. It should be indicated that in all models, speed increase does not have a considerable effect on geographic uniform distribution of nodes.

The order of results in Fig. 4(b) is similar to those of Fig. 4(a), but with an increase in the number of nodes, Geographical uniform distribution metric value reduces.

### 6.2.2 Node movement interdependency

Figure 5 shows node movement interdependency in various speed and various number of nodes respectively. Because of group mobility of nodes, RPGM group mobility model has the highest metric value. OM and RW entity mobility models have similar movement pattern. As the number of destinations and movement pathways in OM realistic model is more restricted, this model shows better results. In both RW and OM models, speed increase leads to less node movement interdependency.

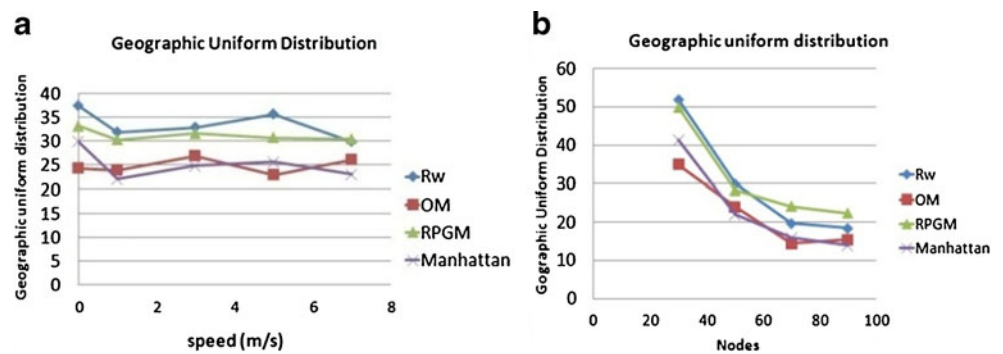
However, the situation for Manhattan mobility model is completely different. In this model, some nodes can continue moving without having any pause time and randomly change their direction at crossroads. Thus, nodes in this model are free more in selecting their movement manner, which causes them to be less dependent in their movement.

Figure 5(b) shows the node movement interdependency metric with various number of nodes. The order of diagrams in Fig. 5(b) is similar to those of Fig. 5(a) and with an increase in the number of nodes, movement interdependency increase too.

**Table 1** Simulation parameters

Mobility and simulation terrain parameters			
Simulator	NS-2	Pause time(in OM,RPGM and RW)	50s
Terrain size	2,000 m <sup>2</sup>	Simulation terrain of Manhattan mobility model	10×10 grid with 0.25Change direct probability
Node speed	0 to 8 m/s	Group size in RPGM	10nodes
Node number	50	Group change probability in RPGM	0.1
Simulation time	500 s	Signal obstruction in OM	no
Over lay structure characteristics		Underlay structure characteristics	
Percentage of online node from beginning simulation	50 %	propagation model	two-ray-ground radio
Number of queries	1,250	MAC protocol	IEEE 802.11b modelled on Cisco AiroNet350
Number of files	30	length of interface queue	30 packets
Channel error rate	0.05	Routing protocol	AODV
		Transport layer	UDP protocol

**Fig. 4** Geographic uniform distribution **a** with various speeds **b** with various number of nodes



### 6.3 Validating and evaluating underlay structure metrics

#### 6.3.1 Average link duration

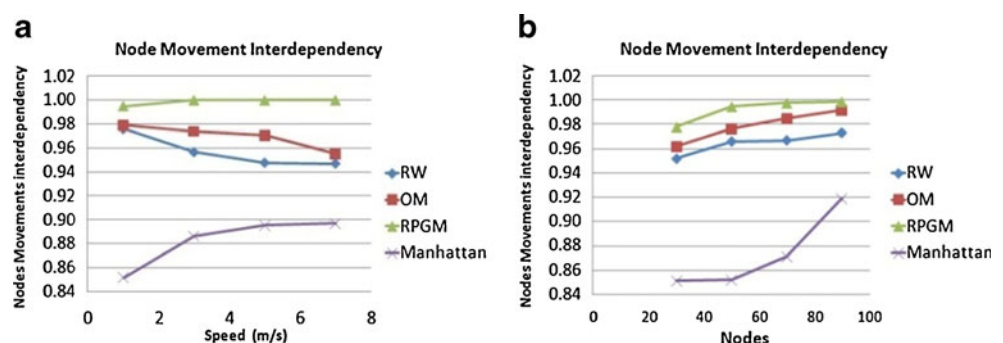
Compared to entity mobility models, the average link duration in RPGM group mobility model is higher (Fig. 6(a)). Considering node group movement, this duration is expected. Among entity mobility models, OM model represented better results. Because, compared to Manhattan and RW mobility model, the number movement pathways are more restricted. Nodes have to move through those pathways. So, during simulation, more links remain connected.

Regarding Fig. 5(a), average link duration in Manhattan model is lower. Because, firstly, it has more movement pathways compared to OM model and secondly, there is no pause time and nodes are always mobile. Speed increase results in lower average link duration. As shown in Fig. 6(b), increasing the number of nodes has no effect on link duration.

#### 6.3.2 Average link changes

The Average Link Changes for different mobility models in various speeds has been shown in Fig. 7(a). The order of results is opposite of Fig. 6(a). In other words, the more link changes, the lower average link duration will be. As it can be seen in Fig. 7(a), the least link changes are related to RPGM group mobility model and then related to OM, RW and Manhattan entity mobility models. The analyses of reasons related to results are similar to the description of Fig. 6(a).

**Fig. 5** Node movement interdependency **a** with various speeds **b** with various number of nodes



Here with speed increase, we can observe a considerable increase in the rate of link changes.

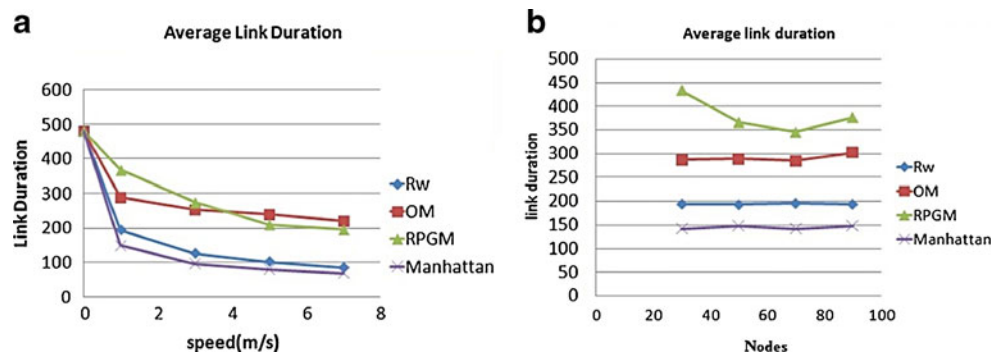
The results of Fig. 7(b) are opposite of Fig. 6(b); increasing the number of nodes has no effect on link changes. In RPGM model, increasing the number of nodes first increases link changes, then decreases link changes. This is because the probability of group membership changes increases, and these changes in groups result in breaking the link that a node has with its previous group members and in creating new links with new group members. However, these changes decrease as the node number reaches 90, considering the size of simulation environment.

#### 6.3.3 Average path availability

Figure 8 shows path availability for different mobility models in various speeds and various number of nodes. Contrary to the expectation, as shown in Fig. 8(a), RW random mobility model had the highest degree of path availability; while in Figs. 6(a) and 7(a), RPGM model had the highest Link duration and lowest Link changes. Because, Unlike RPGM, in RW mobility model there are more number of paths considering geographic uniform distribution, but in RPGM model, paths are established among inside group nodes. The duration of these valid paths in RW mobility model is short.

After this model, there are OM and Manhattan realistic model diagrams. Predefined pathways and non-uniform distribution in those models lead to decline in Path availability. Because of more mobility, this decline is greater in Manhattan mobility model.

**Fig. 6** Average Link Duration **a** with various speeds **b** with various number of nodes



It can be seen that, as shown in Fig. 8(a), speed increasing leads to a decrease in Link duration and Link changes, but it has no effect on Path availability; increasing speed leads to path changes and new paths replace the old ones.

But increasing the number of nodes results in the Path availability increase (Fig. 8(b)). Of course this case is expected, for node density leads to an increase in the number of available and valid paths. A different behavior is observed in RPGM group model. This happens because as the node number increases, the number of groups increases and the probability of network fragmentation increases so that a group of nodes may move to a part of environment in which they cannot connect to other groups.

6.4 Connectivity graph total weight

The lower value of this metric refers that connectivity graph edges are composed of these nodes that are closer to one another and have created a more stable connectivity graph. Among different mobility models, realistic models with predefined pathways present connectivity graph with lower weight (Fig. 9(a)). Because in these models nodes can move freely through pathways and the link length between them is short. Regarding more node movement interdependency in OM model, it has better condition compared to Manhattan mobility model. Among unrealistic models, we expected better result from RPGM group model, because of nodes group mobility and greater interdependency of them. As

there are a number of groups in this model, each of which has movement interdependency, these results are justifiable. It should be mentioned that, in all models, increasing speed leads to an increase in connectivity graph total weight.

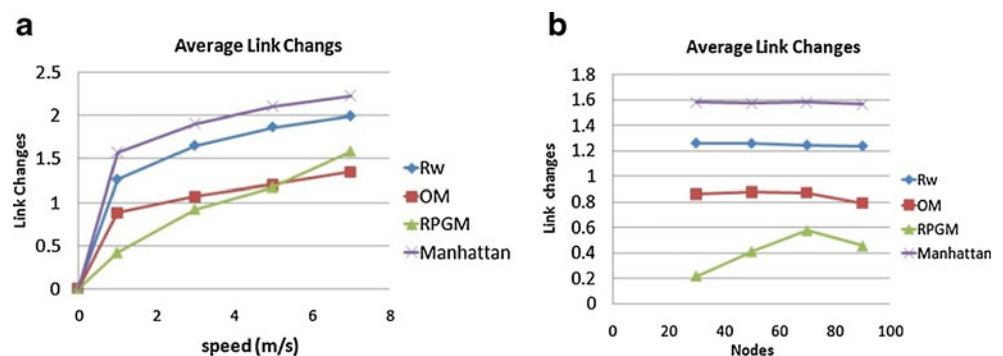
Figure 9(b) shows the metric value of connectivity graph total weight with various number of nodes. When the number of nodes increases, connectivity graph weight increases too, as it creates a graph with greater number of edges and nodes.

6.5 Evaluating and analyzing the relation between mobility and underlay structure

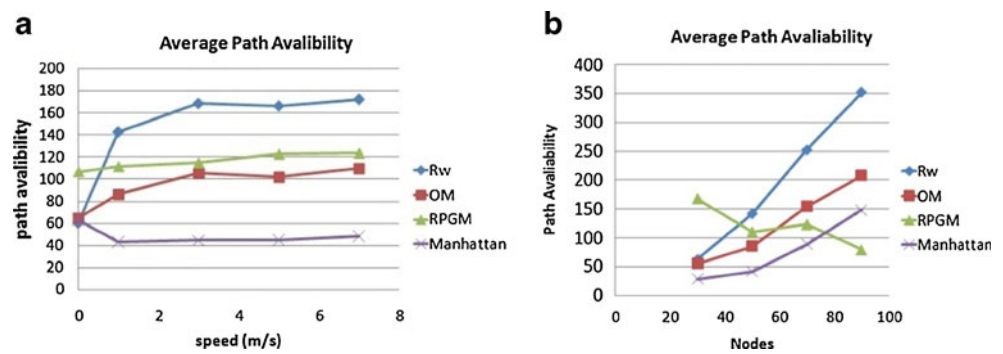
By evaluating the calculated metrics in I and II phases and connectivity graph total weight metric, the following results were obtained:

1. Unrealistic models create a greater geographical uniform distribution for nodes.
2. Node movement interdependency is greater in group models and then in realistic models.
3. In group models and those models which have greater node movement interdependency, Link duration is higher.
4. There is a negative correlation between Link changes and Link duration.
5. Two factors have important roles in Path availability: Geographic uniform distribution and Link duration.

**Fig. 7** Average Link changes **a** with various speeds **b** with various number of nodes



**Fig. 8** Average path availability **a** with various speeds **b** with various number of nodes



## 6.6 Validation and evaluation of overlay structure metrics

### 6.6.1 Hit rate metric

Figure 10 illustrates the hit rate of Chord, Gnutella, and Random walk protocols under the mobility models of RW, OM, RPGM and Manhattan at various speeds.

Like the result of [29], Hit rate for Gnutella protocol has the best value and it is a result of query flooding. For having more Path availability, RW mobility model has the best results among different mobility models (Fig. 10(b)). In both Path availability and Link duration, RPGM model has better results, but after RW mobility model, the highest value of Hit rate is related to OM model. As in RPGM model, some attempts have been made to create connection between nodes of two different groups that may fail. The worst results are related to Manhattan mobility model and regarding the low rate of Path availability and Link duration, these results are expected.

Random Walk protocol has a procedure similar to Gnutella, but the rate of flooded queries is lower. So the order of results in Fig. 10(c) is similar to those of Fig. 10(b) with lower value.

The lowest Hit rate value is related to Chord protocol. There are considerable differences between Hit rate values in zero speed and other speeds (Fig. 10(a)). Therefore, this model is not suitable for mobile nodes. Regarding the structure of overlay in Chord protocol, Link duration and stability, compared to unstructured protocols, has a greater effect on hit rate. Having higher duration and stability (connectivity graph total

weight) OM model has better results. And then there are RW (considering Path availability rate), RPGM and Manhattan mobility model respectively.

For unstructured P2P protocols, speed increase does not have considerable effects on Hit rate. It has been shown in Path availability diagram.

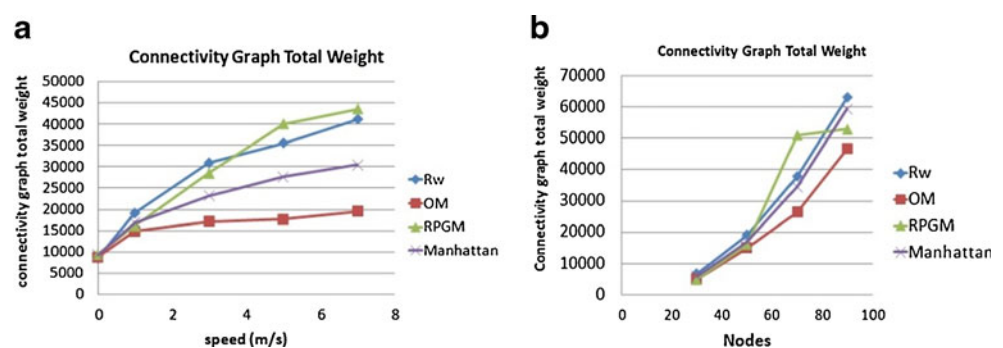
Figure 11 shows that Hit rate value for various number of nodes. Gnutella and Random walk protocols show similar results. The only difference is that Hit rate in Gnutella is higher. The order of results is similar to Fig. 10. With increasing the number of nodes in OM, Manhattan and RW mobility models, first the Hit rate increases and then decreases. The initial increase is because of Path availability increase. Path availability increase in itself is as a result of increasing in the number of nodes. And the decreasing that follows the initial increases is due to congestion and contentions in underlay. Hit rate diagram for RPGM model is exactly similar to Path availability diagram in this model (Fig. 8(b)). In other words, by increasing Path availability in RPGM model, Hit rate increases.

### 6.6.2 Average response time metric

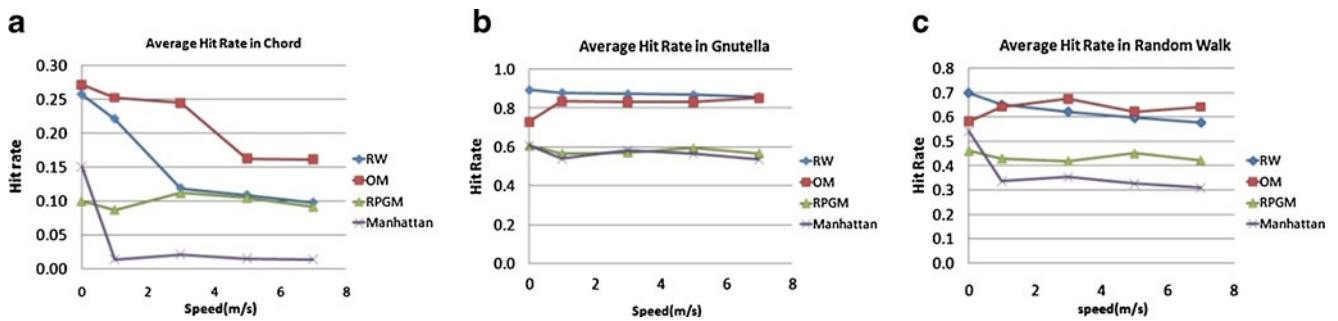
According to the Response time metric definition, the influential factor is time cost of the path between query generator and its response. In this cost, geographic distance (presented by Connectivity graph total weight metric) and the quality of path links (presented by Link duration) are influential.

According to the Fig. 12, it can be observed that, in all content discovery protocols, RPGM model, have the highest

**Fig. 9** Connectivity graph total weight **a** with various speeds **b** with various number of nodes







**Fig. 10** Average hit rate with various speeds in **a** Chord Protocol **b** Gnutella Protocol **c** Random Walk protocol

degree of Link duration and short distance among the present nodes in one group, had the lowest response time. On the other hand it can be concluded that, with respect to Hit rate diagram, this model responds to those queries with source and destination existing in one group.

OM model comes after RPGM model having higher Link duration and lower Connectivity graph total weight compared to RW and Manhattan mobility model. The worst results belong to Manhattan mobility model having the lowest Link duration and highest Connectivity graph total weight. Considering the influence of increasing speed on Link duration and Connectivity graph total weight, speed increase leads to response time increase in all protocols.

When the number of node increases, different behaviors are observed in protocols (Fig. 13). An increase in the number of nodes leads to response time increase in Gnutella and response time decrease in Random walk. This case is due to controlled flooding in Random walk. As increasing the number of nodes creates more suitable paths for walkers to reach the destination. In Gnutella, by increasing the number of nodes, the number of queries which are flooded increases too and the network traffic soars; contention and channel acquisition problem result. In Chord, Response time decrease due to an increase in the number of nodes is contrary to the expectations. According to Fig. 9(b), increasing the number of nodes leads to shorter links in overlay and Response time decrease.

### 6.6.3 Average energy consumption metric

There is a direct relationship between Energy consumption and Hit rate. Regarding flooding behavior and higher Hit rate in Gnutella, Energy consumption in Gnutella is higher than Random walk and Chord Energy consumption (Fig. 14).

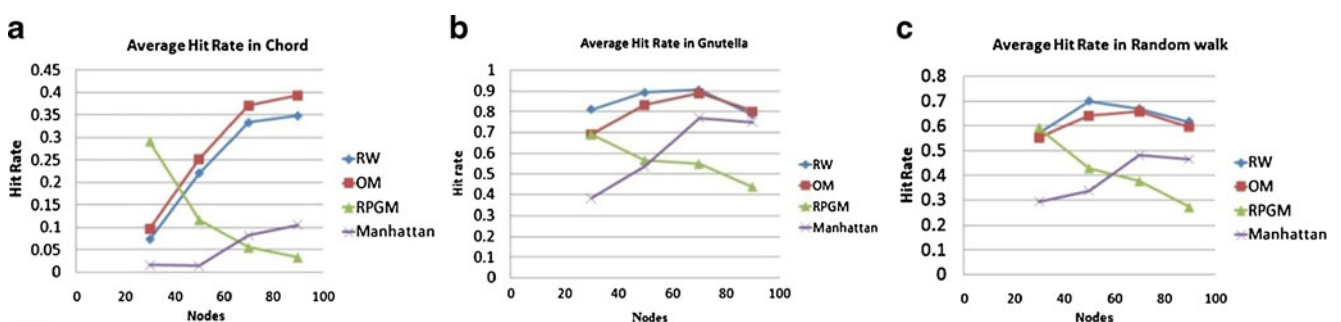
Also in mobility models, there is a correspondence between Energy consumption and Hit rate; of course Link changes are influential too. For instance, Energy consumption in RW and Manhattan mobility model is more than that of OM model and RPGM model respectively. It should be mentioned that speed changes does not have a considerable effect on Energy consumption.

When the number of the node increases, the growth in Energy consumption accelerate more than growth in Hit rate. This is due to contention and traffic increase (Fig. 15).

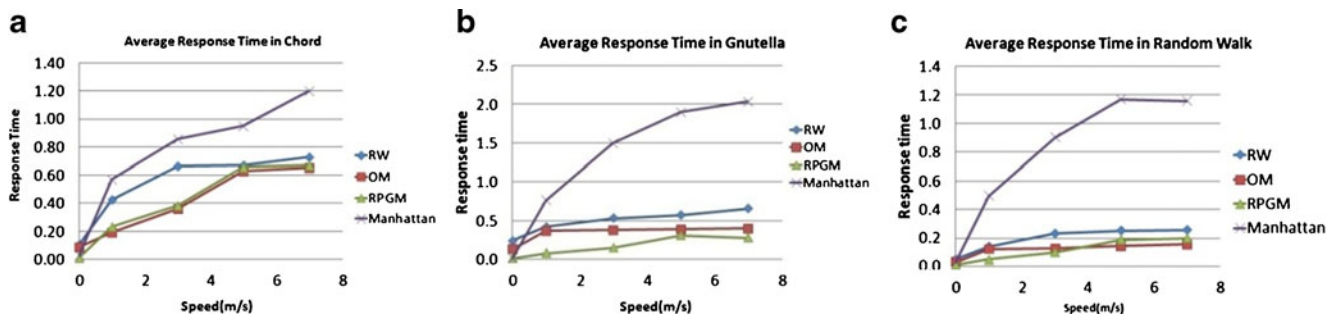
### 6.7 Validating the metrics of average of link matching and average link stability

Figures 16 and 17 illustrate the average Link matching and Link stability metrics, and Link Stability variance for various speed. As increasing the speed does not have a considerable effect on these metrics, the averages of these changes are considered.

The smaller the average link matching value, the more the efficiency will be. Evaluation of this metric in itself cannot describe efficiency. Because this metric has been



**Fig. 11** Average hit rate with various number of nodes in **a** Chord Protocol **b** Gnutella Protocol **c** Random Walk protocol



**Fig. 12** Average response time with various speeds in **a** Chord Protocol **b** Gnutella Protocol **c** Random Walk Protocol

calculated for those links from overlay which have a path in underlay. So link matching metric should be analyzed based on the Stability which shows the establishment of overlay links. Therefore, RPGM and Manhattan mobility models that represent the best results (Fig. 16) do not have a good efficiency in Hit rate. Because only a limited number of links that supported by short path, are connected in overlay and this limited number cannot create a desirable Hit rate. In spite of having shorter links, OM model, compared to RW mobility model, has lower Hit rate because it has lower stability (Fig. 17). It can be concluded that, in MANET environment, stability of overlay links has greater importance.

Theoretically, we expect that with increasing the stability, the efficiency of content discovery protocols increases (Hit rate). So it is observed that the order of results in Fig. 17 for Gnutella and Random walk is similar to those of figures related to Hit rate (Fig. 10(b) and (c)). In RPGM model, considering a higher stability, we have lower Hit rate compared to RW and OM model. It is due to higher variance in stability metric in RPGM model. It means that only those overlay links that belong to same group nodes are stable and it leads to partitioning overlay graph and finally to reducing Hit rate. On the other hand, regarding similar behavior of Random walk and Gnutella in establishing neighborhood, the Link stability in both of them is alike and is more than Link stability in Chord tangibly. Since the way of creating overlay graph in Chord is irrespective of connectivity graph and content search method is based on DHTs, impact of

instability of overlay links in Chord performance is more than unstructured protocols.

The order of results in Fig. 17 for Chord is different from its Hit rate results (Fig. 10(a)). Although RPGM model has a great stability, its variance is high. It means that, in calculating stability average, we had group inside links with good stability and group outside links with bad stability. This fact has a great influence on the efficiency of overlay structure. OM and Manhattan come after RPGM model and their order is similar to Hit rate diagram (Fig. 10(a)).

#### 6.8 Evaluating and analyzing the relation between underlay and overlay structures

With assessing the evaluated metrics in phase 2 and 3 and also link matching and link stability, we can have the following claims. It should be mentioned that the following claims are proven in section 7 by using mathematical analysis.

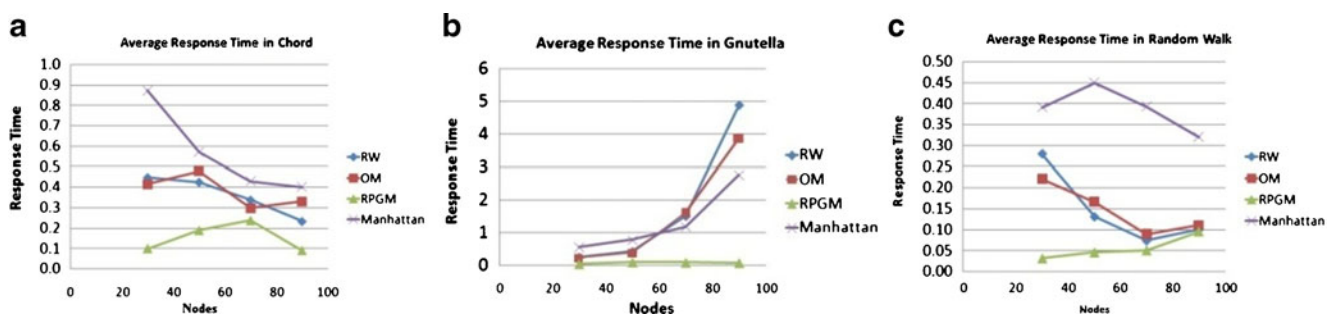
*Claim1:* Path availability and Hit rate parameters are directly related to each other.

*Claim2:* Hit rate and Link stability parameters are directly related to each other.

*Claim3:* Link Duration parameter affects response time.

*Claim4:* Connectivity graph total weight (CGTW) parameter affects response time.

*Claim5:* Link Duration parameter has a considerable effect on energy consumption.



**Fig. 13** Average response time with various number of nodes in **a** Chord Protocol **b** Gnutella Protocol **c** Random Walk Protocol

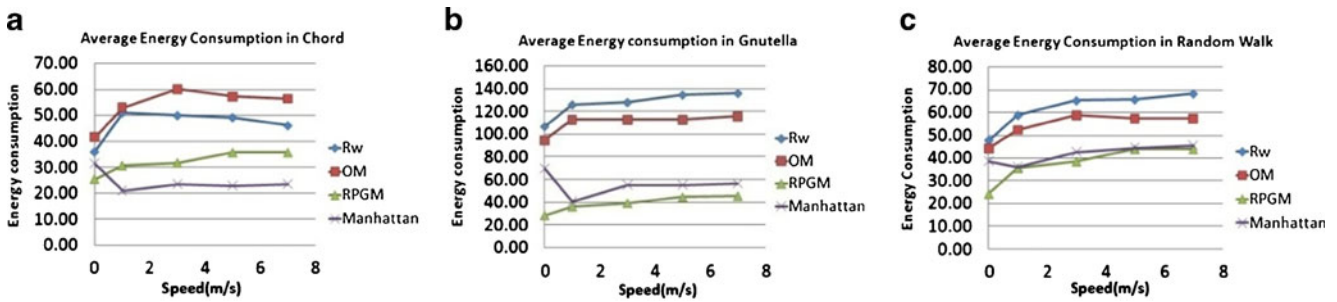


Fig. 14 Average energy consumption with various speeds in a Chord Protocol b Gnutella Protocol c Random walk Protocol

Claim6: Hit rate and Energy consumption parameters are directly related to each other.

7 Mathematical analysis

Mathematical analysis can be the foundation of simulation based claims. The purpose of this section is modeling and analysis of the efficiency of P2P content discovery protocols under MANET by using Generalize Random Graph (GRG) [16]. Regarding the weak results of efficiency of Chord protocol under MANET which were obtained in simulation, we just focus on modeling unstructured protocol here.

Many mathematical models have been presented for unstructured content discovery protocols [37–40]. Random graph is one of these models that has been used flooding modeling in [41]. GRG is a more effective strategy in [42] that has been used for modeling and improving the efficiency of Gnutella. In [43], the same strategy has also been used for modeling flooding search in P2P network and a controlled flooding strategy for its improving has been proposed. In this paper, for modeling P2P content discovery protocol under MANET, we have modeled each snapshot of overlay graph with a GRG.

In GRG, for degree distribution of each randomly chosen vertex  $v$  generating function  $G_0(x) = \sum_{k=0}^{\infty} P_k x^k$  has been used in which  $P_k$  is the probability that a randomly chosen vertex has exactly degree  $k$ . For distribution of the vertex

degree as the first neighbor of  $v$  (the vertex which is located at the end of a random output edge from  $v$ ), generating function  $G_1(x) = \frac{G_0'(x)}{G_0'(1)}$  is used in which  $G_0'(x)$  is derivation of  $G_0(x)$  and  $G_0'(1)$  is the average vertex degree of a randomly chosen node. So generating function of the number of second neighbors of a randomly chosen node like  $v$  equals  $\sum_{k=0}^{\infty} P_k [G_1(x)]^k$  or equals  $G_0(G_1(x))$ . And also generating function of the number of third neighbors of a randomly chosen vertex that has at least a second neighbor is shown by  $G_0(G_1(G_1(x)))$  and so on.

We define a function  $g : N \rightarrow [0, 1]$  representing the query forwarding probability as a function of the query distance from the originator. This function is as follows:

$$g(k) = R(k) \times (F(k)|R(k)) \tag{18}$$

$R(k)$  is query receiving probability in  $k$  distance from originator and  $F(k)$  is the forwarding probability.

The probability that the query originator transmits search message to  $n$  number of its first neighbors is defined by  $q_n = \sum_{h=n}^{\infty} P_h \binom{h}{n} g(0)^n (1 - g(0))^{h-n}$  (we assume  $g(0)$  for the query originator itself).

If  $h$  is the number of first neighbors of query originator,  $h=n$  in Gnutella protocol. But in Random Walk protocol  $n < h$ . Generating function of  $q_n$  probability distribution is as follows:

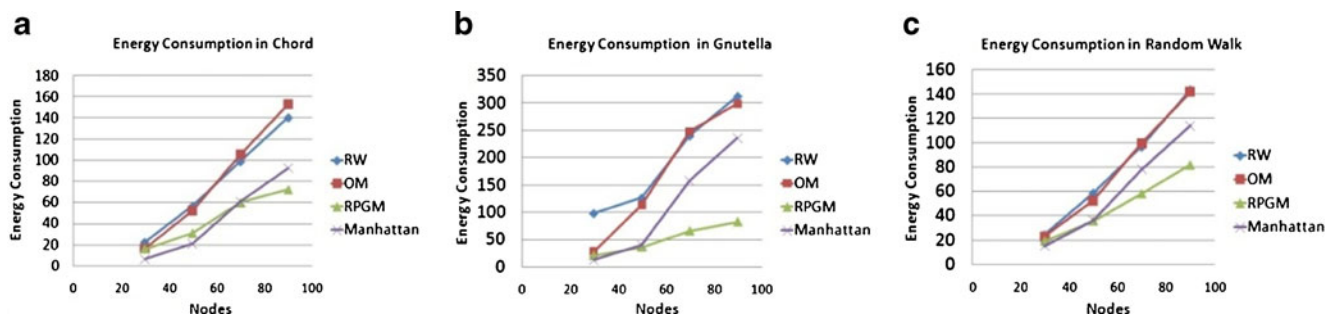
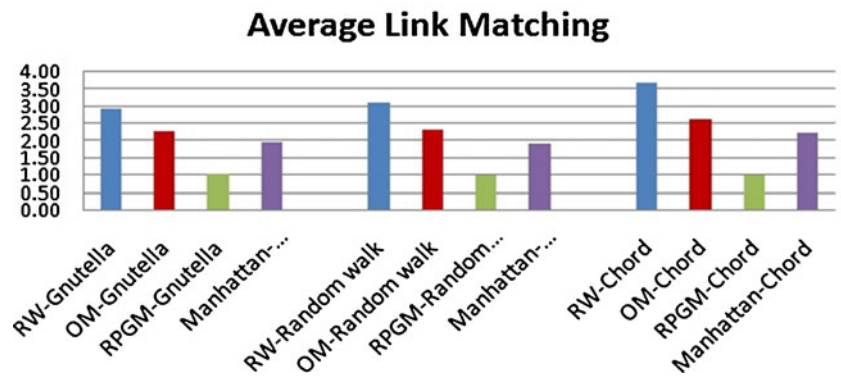


Fig. 15 Average energy consumption with various number of nodes in a Chord Protocol b Gnutella Protocol c Random walk Protocol

**Fig. 16** Average link matching for all P2P content discovery protocols



$$Q_1(x, g) = \sum_{n=0}^{\infty} q_n x^n = \sum_{n=0}^{\infty} \left[ \sum_{h=n}^{\infty} P_h \binom{h}{n} g(0)^n (1 - g(0))^{h-n} \right] x^n$$

$$= \sum_{h=0}^{\infty} P_h (x g(0) + 1 - g(0))^h = G_0(1 + g(0)(x - 1))$$

$Q_1(x, g)$  is generating function of the number of first neighbors of query originator that receive the query message. Generating function of the number of second neighbors of the query originator that receive the query message equals:

$$Q_2(x, g) = \sum_{h=0}^{\infty} q_h [Q_1^{(1)}(x, g)]^h = Q_1(Q_1^{(1)}(x, g), g) \quad (19)$$

In which  $Q_1^{(1)}(x, g) = G_1(1 + g(1)(x - 1))$ . In Random Walk protocol  $q_h$  in  $Q_2(x, g)$  equals  $q_1$ . Because, the first neighbor of query originator on, send query message only to one of its neighbors. So  $Q_h(x, g)$  is generating function of the number of peers at  $h$  distance from the query originator that receive the query message.

$$Q_{h+1}(x, g) = Q_h(Q_1^{(h)}(x, g), g) \quad h \geq 1$$

Where  $Q_1^{(h)}(x, g) = G_1(1 + g(h)(x - 1))$ . Finally, generating function of the total number of query originator

neighbors that receive a copy of query message out to distance  $ttl$  is given by

$$Q(x, g, ttl) = \prod_{m=1}^{ttl} Q_m(x, g) \quad (20)$$

Therefore, the average number of total messages sent for a search equals

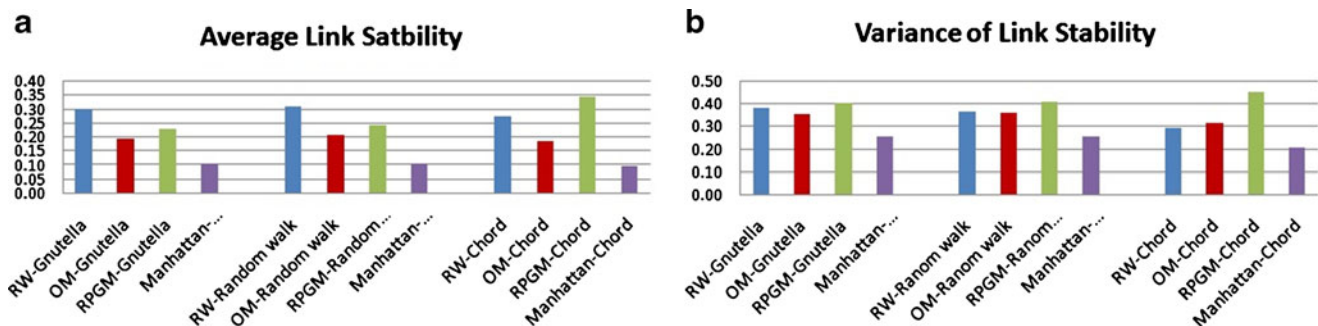
$$\bar{m} = Q'(1, g, ttl) \quad (21)$$

*Definition 1* the average Energy consumption in the network equals:

$$EC = 2\bar{m}c \quad (22)$$

In which  $c$  is a constant factor that equals the average of consumed power for transmitting, receiving and idle modes of a peer. The probability that  $n$  number of the first neighbor of a randomly chosen query originator which receive query message and have a copy of resource  $f$  (response of query) is given by

$$p(n, \gamma) = \sum_{h=n}^{\infty} q_h \binom{h}{n} \gamma^n (1 - \gamma)^{h-n}$$



**Fig. 17** a Average link stability bVariance of link stability for all P2P content discovery protocols



$\gamma$  is the probability of having a copy of requested resource for each randomly chosen peer. Probability generating function  $p(n, \gamma)$  is defined as follows

$$\begin{aligned} H_1(x, g, \gamma) &= \sum_{n=0}^{\infty} p(n, \gamma)x^n \\ &= \sum_{n=0}^{\infty} \left[ \sum_{h=n}^{\infty} q_h \binom{h}{n} \gamma^n (1 - \gamma)^{h-n} \right] x^n \\ &= Q_1(1 + \gamma(x - 1), g) \end{aligned}$$

Similar to Eq. (19) generating function of the number of second neighbors of a randomly chosen query originator that receive a query and have copy of requested resource equals  $H_2(x, g, \gamma) = Q_2(1 + \gamma(x - 1), g)$  and in general

$$H_h(x, g, \gamma) = Q_h(1 + \gamma(x - 1), g) \tag{23}$$

Eventually, generating function for the total number of neighbors of a randomly chosen query originator out to distance  $ttl$  that receives a query message and posses a copy of the requested resource  $f$  is given by

$$H(x, g, \gamma, ttl) = \prod_{m=1}^{ttl} H_m(x, g, \gamma) \tag{24}$$

*Definition 2* the probability that a query is successful (Hit Rate) equals:

$$P_{hit} = 1 - H(0, g, \gamma, ttl) \tag{25}$$

The probability that  $j$  positive responses arrive to a randomly chosen query originator from its first neighbors within  $t$  seconds is given by  $d_1(j, \gamma, t) =$

$\sum_{m=j}^{\infty} p(m, \gamma) \binom{m}{j} (D_1(t))^j (1 - D_1(t))^{m-j}$ .  $D_1(t)$  is the probability a positive reply from its first neighbors is transmitted to a query originator. Generating function of probability distribution  $d_1(j, \gamma, t)$  is as follows

$$\begin{aligned} T_1(x, g, \gamma, t) &= \sum_{j=0}^{\infty} d_1(j, \gamma, t)x^j \\ &= \sum_{j=0}^{\infty} \left[ \sum_{m=j}^{\infty} p(m, \gamma) \binom{m}{j} (D_1(t))^j (1 - D_1(t))^{m-j} \right] x^j \\ &= H_1(1 + (x - 1)D_1(t), g, \gamma) \end{aligned}$$

Generally, generating function of received responses from neighbors out to distance  $h$  equals  $T_h(x, g, \gamma, t) = H_h(1 + (x - 1)D_h(t), g, \gamma)$ . Finally generating function for probability distribution of  $j$  positive responses from all

neighbors of a randomly chosen query originator out to distance  $ttl$  and within  $t$  second is given by

$$T(x, g, \gamma, t) = \prod_{m=1}^{ttl} T_m(x, g, \gamma, t) \tag{26}$$

*Definition 3* the average Response time to a query equals

$$RT = T'(1, g, \gamma, t) \tag{27}$$

Definition 1,2 and 3 describe evaluation metrics for p2p content discovery protocol by using GRG. Claimed relations among different metrics related to analysis framework phases which were revealed by simulation will be supported and proved by the proposed mathematical model.

The equivalent of function  $g(k)$  that was defined in (18) can be given in the following manner.

$$\begin{aligned} g(k) &\equiv g(s, a) \text{ where } k = dist(s, a) \\ g(s, a) &= R(s, a) \times (F(s, a)|R(s, a)) \end{aligned} \tag{28}$$

Where  $s$  is query originator,  $\alpha$  is the neighbor of  $s$  at distance  $k$  that has received query message, and  $dist(s,a)$  is the length of the shortest path between  $s$  and  $a$ .

In order to prove the claimed relation path availability with other metrics, the function  $R(s,a)$  can be defined as follows:

$$R(s, a) = \frac{PA(s, a)}{T} \tag{29}$$

$PA(s,a)$  is obtained from Eq. (9) and it the same path availability between  $a$  and  $s$  and  $T$  is the number of snapshots. In fact  $R(s,a)$  is the availability probability of the path between two peers  $a$  and  $s$ . In other words, if there is a path between them in all snapshots  $R(s,a)=1$ . Therefore  $g(s,a)$  is a function of path availability and considering Eq. (25), Hit rate is a function of  $g(k)$ . So it is proved that there is a direct relation between Hit rate and path availability (*claim1*).

In the same manner, regarding calculation formula for stability, (Eqs. 16 and 17) this parameter is directly related to path availability. As the direct relation between Hit rate and path availability has been proved, it is concluded that Hit rate and stability are directly related and *claim2* is also proved.

In order to prove that link duration is related to other metrics, we assume there is a path between  $a$  and  $s$  that is called  $path_{s-a}$ . Activation probability of such a path equals multiplication of activation probability by each one of the links on the same path. If we call the activation probability of a link as  $p(L) = \frac{LD(i,j)}{T}$  ( $i, j \in L$  in which  $LD(i,j)$  is the link duration between peers  $i$  and  $j$  in underlay (Eq. 7)

and  $T$  is the number of simulation snapshots. Therefore we have:

$$PA(s, a) = P(\text{path}_{s-a}) = \prod_{\substack{\forall L \in \text{path}_{s-a} \text{ and} \\ L \in \text{Underlay}}} \frac{LD(i, j)}{T} (i, j) \in L$$

$$p(L) = \prod_{\substack{\forall L \in \text{path}_{s-a} \text{ and} \\ L \in \text{Underlay}}} \frac{LD(i, j)}{T} (i, j) \in L \quad (30)$$

Based on Eqs. (29) and (30), we can conclude that the probability of receiving a query message from originator ( $R(s, a)$ ) is directly related to link duration metric. It means that Link duration is also related to  $g(k)$  function directly. Therefore, regarding Eqs. (27) and (22), it can be concluded that link duration affects Response Time and Energy Consumption and it can be regarded as the proof of *claim3* and *claim5*. Considering the reverse relation of Link Change and Link Duration, it is not deemed necessary to prove its relation with other parameters.

For proving *claim4* based on the proposed mathematical model, it can be said that the total weight of connectivity graph, with regard to Eqs. 11 and 12, is directly related to geographical distance between two nodes. On the other hand, according to Friis equation (Eq. 31) [44], there is a reverse relation between the power of receiving signals in nodes and their geographical distance toward one another. Therefore, the less distance neighbor nodes are from one another, or the less connectivity graph total weight is, the more probability of Link Duration among them will be. According to *claim3*, increasing Link Duration will decrease Response time. So CGTW decrease will reduce Response Time.

$$P_{\text{recv}}(d) = P_{\text{recv}}(d_0) \cdot \left(\frac{d_0}{d}\right)^r \quad (31)$$

In order to prove the relation between Hit rate and Energy Consumption, again we consider Eq. (25) that represent hit probability. Hit rate increase means rising function  $H(x, g, \gamma, \text{ttl})$  for  $x > 0$  which express the number of query receiving peers and posses a copy of resource. Where function  $H$  is calculated based on function  $Q$  (Eq. 23), so  $Q$  also increases. Therefore, it result in increasing the number of transmitted messages in the network and eventually, transmitting more messages needs more Energy Consumption. So *claim6* is proved.

## 8 Conclusion

This article aims at analyzing the effect of mobility models on the performance of peer to peer content discovery protocols on MANET. To achieve this aim, an analysis framework has been

presented. Analyzing the results of different mobility models, we find out that those movement pattern which have established more uniform distribution of nodes in simulated environment, provide better efficiency. On the other hand movement limitations such as predefined pathways or group movement, result in efficiency reduction of target network.

Mobility, with any pattern, has a direct effect on underlay structure and its metrics such as Link changes, Link Duration, and Path availability metrics. But increasing movement speed does not have a considerable effect on Path availability. On the other hand, Path availability is the most important factor on which efficiency of content discovery protocols depend. So in implementation of efficient overlay networks, mobility is not considered an obstacle.

Structured protocols do not have good efficiency in MANET environment and mobility has extremely a negative effect on efficiency. This issue is related to the method of choosing neighbor and establishing overlay in these kinds of protocols. In these protocols, performance is completely related to stability and optimality of overlay. As this issue cannot be guaranteed in MANET, with disconnecting even one link in overlay, the whole search process fails. While in unstructured protocols, disconnection of one overlay link, these are other, alternative paths to be replaced by that link. It leads to higher performance of unstructured protocols which are implemented in MANET environment.

There are other factors which can improve the efficiency of these protocols, in the previous researches [34], the factor of “topology awareness” was presented. This factor was introduced for formation of overlay whose links are supported by shorter paths in underlay structure. This factor is suitable for networks with stable topology or network low change rate. But in this article, the average link stability metric of overlay was presented that had the greatest effect on the performance of content discovery protocols. This metric is applied for establishing a more stable overlay that leads to efficiency increase.

**Acknowledgement** This research was financially supported by Iran Telecommunication Research Center (ITRC), as a Ph.D. thesis support program.

## References

1. Haas ZJ, Deng J, Liang B, Papadimitratos P, Sajama S (2002) Wireless ad hoc networks. Wiley Encyclopedia of Telecommunications
2. Schollmeier R (2001) A definition of peer to peer networking for classification of peer-to-peer architecture and applications. In Proceedings of the First International Conference on Peer-to-Peer Computing (P2P '01). IEEE Computer Society, inkoping, Sweden, August 27–29, 2001
3. Androutsellis-Theotokis S, Spinellis D (2004) A survey of peer-to-peer content distribution technologies. ACM Comput Surv 36(4):335–371. doi:10.1145/1041680.1041681

4. Oliveira LB, Siqueira IG, Loureiro AAF (2005) On the performance of ad hoc routing protocols under a peer-to-peer application. *J Parallel Distr Comput* 65(11):1337–1347. doi:10.1016/j.jpdc.2005.05.023
5. Franciscani FP, Vasconcelos MA, Couto RP, Loureiro AAF (2005) (re)configuration algorithms for peer-to-peer over ad hoc networks. *J Parallel Distr Comput* 65(2):234–245. doi:10.1016/j.jpdc.2004.09.007
6. Hu YC, Das SM, Pucha H (2003) Exploiting the synergy between peer-to-peer and mobile ad hoc networks. In *HotOS-IX: Ninth Workshop on Hot Topics in Operating Systems*, pp 37–42
7. Ding G, Bhargava B (2004) Peer-to-peer file-sharing over mobile ad hoc networks. In *2th IEEE Annual Conference on Pervasive Computing and Communications Workshops*, pp 104–108
8. Breslau L, Estrin D, Fall K, Heidemann J, Helmy A, Huang P, McCanne S, Varadhan K, Xu Y, Yu H (2000) Advances in network simulation. *IEEE Comput* 33(5):59–67. doi:10.1109/2.841785
9. Jardosh AP, Belding-Royer EM, Almeroth KC, Suri S (2005) Real-world environment models for mobile network evaluation. *IEEE Journal on Special Areas Commun* 23(3):622–632. doi:10.1109/JSAC.2004.842561
10. Bai F, Sadagopan N, Helmy A (2003) The IMPORTANT framework for analyzing the Impact of mobility on performance of routing protocols for Adhoc Networks, Elsevier. *Ad Hoc Network* 1(4):383–403. doi:10.1016/S1570-8705(03)00040-4
11. Hong X, Gerla M, Pei G, Chiang CC (1999) A group mobility model for ad hoc wireless networks. In *Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems (MSWiM '99)*. ACM, New York, NY, USA, 53–60. doi:10.1145/313237.313248
12. Stoica I, Morris R, Liben-Nowell D, Karger DR, Kaashoek MF, Dabek F, Balakrishnan H (2003) Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Trans Networking* 11(1):17–32. doi:10.1109/TNET.2002.808407
13. Gkantsidis C, Mihail M, Saberi A (2006) Random walks in peer-to-peer networks: algorithms and evaluation. *Perform Eval* 63(3):241–263. doi:10.1016/j.peva.2005.01.002
14. Boukerche A, Zarrad A, Araujo R (2006) Smart Gnutella overlay formation for collaborative virtual environments over mobile ad hoc networks. In *Proceedings of the 10th IEEE international symposium on Distributed Simulation and Real-Time Applications (DS-RT '06)*. IEEE Computer Society, Washington, DC, USA, 143–156. doi:10.1109/DS-RT.2006.6
15. Zheng Q, Hong X, Ray S, (2004) Recent advances in mobility modeling for mobile ad hoc network Research. In *Proceedings of the 42nd annual Southeast regional conference (ACM-SE 42)*. ACM, New York, NY, USA, 70–75. doi:10.1145/986537.986554
16. Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64:026118. doi:10.1103/PhysRevE.64.026118
17. Johnson DB, Maltz DA, Broch J (2001) DSR: the dynamic source routing protocol for multi-hop wireless ad hoc networks. In *Ad Hoc Networking*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA 139–172
18. Perkins CE, Bhagwat P (1994) Highly dynamic destination sequenced distance vector routing (DSDV) for mobile computers. In *Proceedings of the conference on Communications architectures, protocols and applications (SIGCOMM '94)*. ACM, New York, NY, USA, 234–244. doi:10.1145/190314.190336
19. Perkins C E, Royer E M (1999) Ad hoc on-demand distance vector routing. In: *2nd IEEE Workshop on Mobile Computing System and Applications*, 90–100
20. Broch J, Maltz DA, Johnson DB, Hu YC, Jetcheva J (1998) A performance comparison of multi-hop wireless ad hoc network routing protocols. In *Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ACM, October 1998
21. Hong X, Kwon T, Gerla M, Gu D, Pei G (2001) A mobility framework for ad hoc wireless networks. In *Proceedings of the Second International Conference on Mobile Data Management (MDM '01)*, Springer-Verlag, London, UK, 185–196.
22. da Hora DN, Macedo DF, Nogueira JMS, Pujolle G (2007) Optimizing peer-to-peer content discovery over wireless mobile ad hoc networks. In: *The Ninth IFIP/IEEE International Conference on Mobile and Wireless Communications Networks (MWCN)*, 6–10
23. Borg J (2003) A comparative study of ad hoc & peer to peer networks, Master's thesis, University College London
24. Gerla M, Lindemann C, Rowstron A (eds.) (2005) *Peer-to-Peer Mobile Ad Hoc Networks – New Research Issues*, no. 05152 in Dagstuhl Seminar Proceedings
25. Schollmeier R, Gruber Niethammer F (2003) Protocol for peer-to-peer networking in mobile environments. In *12th International Conference on Computer Communications and Networks (ICCCN 2003)*, 121–127
26. Lee U, Park JS, Yeh J, Pau G, Gerla M (2006) Code torrent: content distribution using network coding in vanet. In *Proceedings of the 1st international workshop on Decentralized resource sharing in mobile computing and networking (MobiShare '06)*. ACM, New York, NY, USA, 1–5. doi:10.1145/1161252.1161254
27. Mecella M, Angelaccio M, Krek A, Catarci T, Buttarazzi B, Dustdar S (2006) Workpad: an adaptive peer-to-peer software infrastructure for supporting collaborative work of human operators in emergency/disaster scenarios. In *CTS'06: Proceedings of the International Symposium on Collaborative Technologies and Systems*, IEEE Computer Society, Washington, DC, USA, 173–180
28. Oliveira LB, Siqueira I, Macedo DF, Loureiro AA, Nogueira JM (2005) Evaluation of peer-to-peer network content discovery techniques over mobile ad hoc networks. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 51–56
29. da Hora DN, Macedo DF, Oliveira LB, Siqueira IG, Loureiro AAF, Nogueira JM, Pujolle G (2009) Enhancing peer-to-peer content discovery techniques over mobile ad hoc networks. *Comput Commun* 32:1445–1459. doi:10.1016/j.comcom.2009.04.005
30. Romoozi M, Babaei H, Fathi M (2009) A cluster-based mobility model for intelligent nodes in Ad hoc Networks, ICCSA, LNCS 5592, Springer-Verlag Berlin Heidelberg, 804–817
31. de Berg M, van Kreveld M, Overmars M, Schwarzkopf O (2000) *Computational geometry: algorithms and applications*. Springer Verlag
32. Eichler S, Schroth C, Eberspächer J (2006) Car-to-car communication. In *Proceedings of the VDE-Kongress - Innovations for Europe, 2006Aachen, Germany*
33. Rodgers JL, Nicewander AW (1988) Thirteen ways to look at the correlation coefficient. *American Statistician* 42(1):59–66. doi:10.2307/2685263
34. Rostami H, Habibi J (2007) Topology awareness of P2P overlay networks. *J Concurrency Comput Pract Ex* 19(7):999–1021. doi:10.1002/cpe.v19:7
35. Aschenbruck N, Ernst R, Gerhards-Padilla E, Schwamborn M (2010) BonnMotion: a mobility scenario generation and analysis tool. In *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools '10)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, article51,10 pages. doi:10.4108/ICST.SIMUTOOLS2010.8684
36. Held G (2004) Focus on the Cisco Aironet 350 wireless access point. *Int J Netw Manag* 14(1):3–7. doi:10.1002/nem.487
37. Wu B, Kshemkalyani AD (2008) Analysis models for unguided search in unstructured P2P networks. *Int J Ad Hoc Ubiquit Comput* 3(4):255–263. doi:10.1504/IJAHUC.2008.018911
38. Tang X, Xu J, Lee WC (2008) Analysis of TTL-based consistency in unstructured peer-to-peer networks. *IEEE Trans Parallel Distributed Systems* 19(12):1683–1694. doi:10.1109/TPDS.2008.44v



39. Gaeta R, Sereno M (2011) Generalized probabilistic flooding in unstructured peer-to-peer networks. *IEEE Trans Parallel Distr Syst* 22(12):2055–2062. doi:10.1109/TPDS.2011.82
40. Leontiadis E, Dimakopoulos VV, Pitoura E (2004) Cache updates in a peer-to-peer network of mobile agents. In: *Proceedings of the Fourth International Conference on Peer-to-Peer Computing (P2P '04)*, IEEE Computer Society, Washington, DC, USA, 10–17. doi:10.1109/P2P.2004.12
41. Kant K (2003) An analytic model for peer to peer file sharing networks. In *Proceedings of the International Communications Conference*, Anchorage, AL, USA, May, 2003
42. Adamic LA, Lukose RM, Punyani AR, Huberman BA (2001) Search in power-law networks. *Phys Rev E* 64:46135–46143
43. Gaeta R, Balbo G, Bruell S, Gribaudo M, Sereno M (2005) A simple analytical framework to analyze search strategies in large-scale peer-to-peer networks. *Perform Eval* 62(1):1–16. doi:10.1016/j.peva.2005.07.023
44. Pozar DM (1998) *Microwave engineering*, 2nd edn. Wiley



**Hamideh Babaei** is currently PhD student at Science & research branch of Islamic Azad University in Iran. She received Bs in software engineering from the University of Kashan at 2003, and his MS in computer science at 2005 in Iran. She is a faculty member of Islamic Azad University (Naragh branch). She has taught in the areas of Wireless Networks, Ad hoc and Sensor Networks and her research interests include Semantic Web, Information Retrieval, and recent research focusing on the Mo-

bility model and routing protocol in ad hoc networks. She has published several articles in international conferences and LNCS series. Email: hamidehbabaei@yahoo.com



**Mahmood Fathy** received the B.S. degree in electronics from Iran University of Science and Technology, Tehran, Iran, in 1985, the M.S. degree in computer architecture from Bradford University, West Yorkshire, U.K., in 1987, and the Ph.D. degree in image processing computer architecture from the University of Manchester Institute of Science and Technology, Manchester, U.K., in 1991. Since 1991, he has been an Associate Professor with the Department of

Computer Engineering, Iran University of Science and Technology. His research interests include the quality of service in computer net-

works, including video and image transmission over Internet, the applications of vehicular ad hoc networks in intelligent transportation systems, and real-time image processing, with particular interest in traffic engineering. Email: mahfathy@iust.ac.ir



**Reza Berangi** received the B.Sc. degree in telecommunications and the M.Sc. degree in electronic engineering from Iran University of Science and Technology (IUST), Tehran, Iran, in 1985 and 1989 respectively, and the Ph.D. degree in mobile communications from Victoria University, Melbourne, Australia, in 1998. From 1986 to 1988, he worked with the Nasr Research and Development Centre, Tehran, Iran on signal processing hardware and software design. From 1997

to 1999, he was with the Mobile Communications and Signal Processing Group at VUT, working on smart antenna mobile base stations. From 1999 to 2001, He was with the Australian Cooperative Telecommunication Research Centre (ATCRC), working on intelligent base stations for the 3 G mobile communication systems. Since 2001, Dr. Berangi has been as an assistant professor at the computer department of IUST. In 2010 his visited the Center for Telecommunications and informatics, Victoria University, Australia and did 6 month research on cognitive radio networks. His research interest is signal processing for mobile communications including adaptive algorithms, smart antenna, and cognitive radio networks. Email: rberangi@iust.ac.ir



**Morteza Romoozi** is currently PhD student at Science & research branch of Islamic Azad University in Iran. He received Bs in software engineering from the University of Kashan at 2003, and his MS in computer science at 2006 in Iran. He is a faculty member of Islamic Azad University (Kashan branch). He has taught in the areas of Wireless Networks, Ad hoc and Sensor Networks and his research interests include Semantic Web, Information Retrieval, and recent research focusing on the Mobility

model and routing protocol in adhoc networks. He has published several articles in international conferences and LNCS series. Email: mromoozi@gmail.com