Fuzzy Hindi WordNet and Word Sense Disambiguation Using Fuzzy Graph Connectivity Measures

AMITA JAIN, Ambedkar Institute of Advanced Communication Technologies and Research
D. K. LOBIYAL, Jawaharlal Nehru University

8

In this article, we propose Fuzzy Hindi WordNet, which is an extended version of Hindi WordNet. The proposed idea of fuzzy relations and their role in modeling Fuzzy Hindi WordNet is explained. We mathematically define fuzzy relations and the composition of these fuzzy relations for this extended version. We show that the concept of composition of fuzzy relations can be used to infer a relation between two words that otherwise are not directly related in Hindi WordNet. Then we propose fuzzy graph connectivity measures that include both local and global measures. These measures are used in determining the significance of a concept (which is represented as a vertex in the fuzzy graph) in a specific context. Finally, we show how these extended measures solve the problem of word sense disambiguation (WSD) effectively, which is useful in many natural language processing applications to improve their performance. Experiments on standard sense tagged corpus for WSD show better results when Fuzzy Hindi WordNet is used in place of Hindi WordNet.

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing

General Terms: Algorithms, Design, Languages

Additional Key Words and Phrases: Centrality, fuzzy logic, fuzzy graph, Hindi WordNet, lexicon, word sense disambiguation

ACM Reference Format:
Amita Jain and D. K. Lobiyal. 2015. Fuzzy hindi WordNet and word sense disambiguation using fuzzy graph connectivity measures. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 15, 2, Article 8 (December 2015), 31 pages.
DOI: http://dx.doi.org/10.1145/2790079

1. INTRODUCTION

WordNet [Fellbaum 1998] is an online lexical reference system that has become an important resource for natural language processing (NLP). It has been used in many NLP applications, such as information retrieval, text summarization, machine translation, and question answering [Miller and Fellbaum 2007]. Its design is inspired by the psycholinguistic principle of human lexical memory. It contains nouns, verbs, adjectives, and adverbs, and describes different types of relationships among them. Unlike in a traditional dictionary, in Hindi WordNet the lexical information is organized in terms of word meanings rather than word forms. Word senses in Hindi WordNet are represented rationally by a synonym set (synset), i.e., the set of all words having similar meanings [Narayan et al. 2002]. Hindi WordNet encodes lexical and semantic relations, wherein the former connects pairs of word senses and the latter relates synsets. For

Authors' addresses: A. Jain, Department of CSE, Ambedkar Institute of Advanced Communication Technology and Research, Delhi; D. K. Lobiyal, School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi.

DOI: http://dx.doi.org/10.1145/2790079

instance, there is a semantic relation "meronymy" between आम(āma) and गुठली(guṭhalī) and "holonymy" between गुठली(guṭhalī) and आम(āma). All relations defined in Hindi WordNet are crisp in nature, i.e., the terms/words are either fully related or not related at all. But in many real-life scenarios expressed in natural language, relations between words (terms may interchangeably be used for words) as a matter of degree. In a natural language, some words are related and some are not, and in between there is a gradual transition from not being related to being related. Therefore, it seems more natural to represent an association between words as a matter of degree, i.e., there is a gradual transition from not being related to being related. Therefore, it seems more natural to represent an association between words by a mapping W × W → [0, 1] (with W being the universe of Hindi words), i.e., by a fuzzy relation instead of a traditional relation. For example, लडकी(laḍakī) and औरत(aurata), विद्यालय(vidyālaya) and संस्थान(saṁsthāna) are closely related (since they share approximately similar meaning) and can replace each other in many contexts. But in Hindi WordNet, partial truth is not defined; therefore, these words are considered to be "not related" because लडकी(laḍakī) and औरत(aurata), विद्यालय(vidyālaya) and संस्थान(saṁsthāna) are not exact synonyms. Similar observation can be made for other relations existing between words in Hindi WordNet, like the "has-part" (meronymy/holonymy) relation. In the real world, a concept may partially belong to the other concept, e.g., पेड(pēḍa) and फल(phala). It is easy to observe the fact that every tree doesn't have fruit, but most trees do. Since context plays a significant role in NLP, fuzzification of relations may help to process natural language in a better way. As Hindi WordNet defines only classical relations (also called *crisp relations*), the words in natural language with partial relationships cannot be processed by the Hindi WordNet. Therefore, there is a strong need to extend Hindi WordNet by incorporating fuzzy relations to process the natural language text in a more accurate and efficient manner.

In this article, we propose enriching Hindi WordNet by extending it in two ways. First, we employ the concept of fuzzy graphs to represent Fuzzy Hindi WordNet. This allows a generalization of the relations in Hindi WordNet. The relation can now be given a membership value between [0, 1] rather than simply representing the existence or nonexistence of a relation. Here, the idea of strength of relation plays a significant role. The second and more interesting concept proposed is the composition of different types of fuzzy relations. We also propose fuzzy graph connectivity measures that include both local and global measures. Further, we show how the results of word sense disambiguation (WSD) can be improved by using Fuzzy Hindi WordNet compared to Hindi WordNet.

The rest of this article is organized as follows. In Section 2, the work available in the literature related to fuzzy set theory and NLP are discussed. In Section 3, we describe a review of preliminaries, namely fuzzy set theory and fuzzy graphs, and briefly introduce Hindi WordNet, which was developed at IIT Bombay, India. In Section 4, we propose Fuzzy Hindi WordNet. Section 5 illustrates how WSD can be achieved at a higher level using Fuzzy Hindi WordNet. Section 6 represents the time complexity, experimental setup, and results when Hindi WordNet/Fuzzy Hindi WordNet is used. Finally, in Section 7, we offer our conclusion.

2. RELATED WORK

To incorporate the imprecise and uncertain information in natural language [Bezdek et al. 1986; Chen et al. 2001, 2003; Cock et al. 2005; Diou et al. 2006; Martin 1992], the fuzzy set theory [Zadeh 1965] has been widely applied. In the early 1990s, Martin [1992] presented the concept of space as a network of fuzzy relationships called *fuzzy concept networks* by allowing four types of fuzzy relations between concepts, i.e., positive association, negative association, generalization, and specialization. Further, in the early years of the 21st century, fuzzy concept networks were used by Chen et al. [2001, 2003] to enhance the performance of information retrieval. Since the construction of fuzzy concept networks manually was a tedious task, a method to construct the networks was developed [Chen et al. 2003]. Initially, Rosenfeld et al. [1975] reviewed the basic properties of fuzzy relations and generalized to the case where the underlying set is a fuzzy set. Fuzzy analogues of several basic graph-theoretic concepts, e.g., bridges and trees, are also defined by Rosenfeld et al. Then, Sunitha [2001] proposed the concepts of fuzzy bridges, fuzzy cutnodes, fuzzy trees, blocks, and metrics in fuzzy graphs defined by Rosenfeld et al. [1975]. Sunitha also modified the definition of the complement of a fuzzy graph. More recently, fuzzy graphs have been used by Yager [2010] to represent the relations in social networks. Fuzzy declarative approach is also used by researchers to classify unlabeled short texts [Romero et al. 2012].

Navigli [2009] completed a survey on WSD in which its methods were divided into three subcategories, namely supervised, unsupervised, and knowledge-based methods. In his work, the author mentioned some other approaches, i.e., domain-driven disambiguation. To deal with the ambiguity in natural language, Navigli and Lapata [2010] used the measures of graph connectivity from social network analysis like degree centrality [Freeman 1979], the key player problem [Borgatti 2006], Hypertext Induced Topic Selection (HITS) [Gupta 2006], and PageRank and betweenness centrality [Freeman 1979; Newman 2005]. Navigli and Lapata [2010] also categorized the results into two categories, namely local measures and global measures. In their work, the researchers used WordNet sense inventory [Fellbaum 1998]. This indicates that the graph-based WSD algorithm performs better with more densely connected sense inventories and more incident edges for every node. The performance of these NLP applications highly depends on the sense inventory used (WordNet). Moreover, since the acquisition of lexicons for this was a challenging task, the methods to automatically acquire lexicons were proposed [Widdow and Dorow 2002]. Extension of WordNet was also proposed by Navigli [2005]. However, one of the main obstacles to high-performance WSD was the knowledge acquisition bottleneck. Ponzetto and Navigli [2010] presented a methodology to automatically extend WordNet with large amounts of semantic relations from Wikipedia. In 2010, a work on translating politeness across cultures was done [Kumar and Jha 2010]. In 2013, Sheinman worked on identifying and encoding intensity relations among adjectives in WordNet [Sheinman et al. 2013]. Lack of labeled data is one of the most severe problems facing WSD. Fujita and Fujino [2013] overcome the problem by combining automatic labeled data expansion and semisupervised learning. Recently, researchers worked on automatically resolving ambiguity in query expansion to improve retrieval performance [Jain et al. 2014a].

Researchers at IIT Bombay developed Hindi WordNet, which is widely used for many NLP applications. Hindi WordNet is used in resolving ambiguity in Hindi language [Jain et al. 2013; Sinha et al. 2004; Dwivedi and Rastogi 2008]. Hindi WordNet is also used for Hindi query expansion [Das et al. 2010] and query optimization [Kumar and Mansotra 2012]. IndoWordNet was built by the expansion method from HindiWordNet [Bhattacharyya 2010]. It groups 16 of India's 22 languages. It is also linked to the English synsets. By exploiting ontology in Hindi WordNet, information retrieval is optimized by the researchers [Sharan et al. 2011]. Jain et al. [2014b] used Wikipedia to automatically generate synsets in WordNets for Indian languages. Recently, Jain et al. [2014c] used Hindi WordNet for retrieving Web search results in soft clusters for a Hindi query.

After going through the literature, we find that the researchers have not explored the possibility of partial belongingness (fuzziness) that may exist in Hindi words among themselves. Therefore, in this article, we propose the concept of fuzzy relations in Hindi

WordNet and subsequently apply it to WSD of Hindi text using fuzzy graph connectivity measures proposed by us in Section 5.

3. PRELIMINARIES

In this section, we introduce the preliminary concepts of fuzzy sets, fuzzy graphs, and Hindi WordNet that we use in this work.

3.1. Fuzzy Set Theory and Fuzzy Graph

A mathematical framework to describe fuzzy logic was suggested by Zadeh [1965] in his seminal paper entitled "Fuzzy Sets." A sharp, unambiguous distinction exists between the members and nonmembers of the class represented by the crisp set. However, many of the terms that we commonly use, such as *tall* and *beauty* (called *linguistic variables*), do not exhibit this characteristic. Klir and Folger [1998] call this a *mismatch problem*: the world is gray, but science is black and white. The fuzzy principle proposed by Zadeh is that "everything is a matter of degree." Thus, the membership in a fuzzy set is not a matter of affirmation or denial but rather a matter of degree.

Membership function μ_A (also known as characteristic function) in crisp set A maps the members in universal set X to set {0, 1}. The concept of fuzzy sets is a generalization of the crisp sets [Lee 2005]. In fuzzy sets, each element is mapped to [0, 1] by membership function

$$\mu_A : X \rightarrow [0, 1],$$

where [0, 1] denotes the set of real numbers between 0 and 1 (including 0, 1). Consequently, a fuzzy set is a vague boundary set as compared to a crisp set [Yen and Langari 2005].

A fuzzy graph [Rosenfeld et al. 1975; Sunitha 2001; Zadeh et al. 1975; Mathew and Sunitha 2009; Bhattacharya 1987] defined as $G = (\sigma, \mu)$ is a pair of functions $\sigma : S \rightarrow [0, 1]$ and $\mu : S \times S \rightarrow [0, 1]$, where for all *x, y* in S we have $\mu(x, y) \leq \sigma(x) \cap \sigma(y)$. Any relation $R \subseteq S \times S$ on a set S can be represented as a graph with node set S and arc set R. Similarly, any fuzzy relation $\mu : S \times S \rightarrow [0, 1]$ can be regarded as defining a fuzzy graph, where the arc $(x, y) \in S \times S$ has weight $\mu(x, y) \in [0, 1]$.

3.2. Hindi WordNet

Hindi WordNet is a system for bringing together different lexical and semantic relations between Hindi words. It organizes the lexical information in terms of word meanings and can be termed as a lexicon based on psycholinguistic principles. Hindi WordNet is widely used in many NLP applications [Avneet 2010; Jain et al. 2013; Mishra et al. 2009; Sinha and Siddiqui 2012; Sinha et al. 2004]. In this, for each word there is a synset representing one lexical concept. Synsets are the basic building blocks of Hindi WordNet. The lexicon deals with the content words or open-class category of words. Thus, Hindi WordNet contains the following categories of words: nouns, verbs, adjectives, and adverbs.

Each entry in Hindi WordNet consists of the entries synset, gloss (description of concept), and position in ontology.

3.2.1. Relations in Hindi WordNet. Hindi WordNet is a network of word senses. A word sense node in this network is a synset that is regarded as a basic object in WordNet. Each synset in Hindi WordNet is linked to other synsets through the well-known lexical and semantic relations such as hypernymy, hyponymy, meronymy, troponymy, antonymy, and entailment. Semantic relations exist between synsets, and lexical relations exist between words. These relations serve to organize the lexical knowledge base. In Hindi WordNet, relations are also defined to connect cross parts of speech, namely linkage between nominal and verbal concepts, nominal and adjectival concepts, and adverbial and verbal concepts. For the sake of brevity, other details of Hindi WordNet are omitted. For more clarity introduction to Hindi WordNet at http://www.cfilt.iitb.ac.in/wordnet/webhwn/ can be referred.

Although Hindi WordNet has become an indispensible resource for NLP, the main obstacle to high-performance NLP applications is the knowledge acquisition bottleneck. Next, we describe some of the prominent causes.

Absence of proper expressiveness. Typically, only classical relations are defined, i.e., either two concepts are fully related or not related at all. However, the natural language text does not always contain classical relations due to the inherent vagueness present in the human language. Sometimes the relations are "approximate" in nature. It is therefore necessary to consider these approximate relations as well to fully understand the natural language text. Consider the pairs of words (विद्यालय(vidyālaya), संस्थान(saṁsthāna)), (अलमारी(alamārī), रैक(raika)), and (लडकी(laḍakī), औरत(aurata)): these are not synonyms under the classical lexicon, although the words are approximately similar and interchangeable in many contexts. In this article, we exploit this observation to fuzzify Hindi WordNet.

Missing composition of semantic relationships. The composition of similar/dissimilar relations is not defined in Hindi WordNet. Although Hindi WordNet is a computational lexicon that resembles an associative network, due to undefined composition of relations, presently it is rather a passive data structure that does not support inferences. For instance, (वाहन(vāhana), कार(kāra)) shares a hypernymy–hyponymy relation and (कार(kāra), पहिया(pahiyā)) shares a meronymy–holonymy relation, but the relation between (वाहन(vāhana), पहिया(pahiyā)) is not defined in Hindi WordNet. In the present work, we define the composition of fuzzy relations to extract the exact meaning of real-world concepts such as (वाहन(vāhana), पहिया(pahiyā)) when they are used in natural language sentences.

4. FUZZY HINDI WORDNET

In this section, we fuzzify the relations present in Hindi WordNet. The fuzzy relations can exhibit semantics between the words more effectively, and therefore fuzzification of Hindi WordNet overcomes the limitations of Hindi WordNet discussed in the previous section. The proposed fuzzified Hindi WordNet by incorporating fuzzy relations between the words is referred to as Fuzzy Hindi WordNet in this work. Since it is believed that fuzziness is more natural [Martin 1992], Fuzzy Hindi WordNet can provide more expressiveness than Hindi WordNet.

We render Fuzzy Hindi WordNet as a fuzzy graph where nodes represent concepts (synsets) and edges represent fuzzy relations between concepts. The weight of an edge represents the strength of the relation between two concepts/synsets. The value of strength varies from 0 to 1. More value of strength indicates a closer relationship between the concepts, i.e., synsets. It is clear that a strength value of 1 indicates a classical relation between the words/concepts, whereas a 0 value indicates no relation. It may be noted that since fuzzy logic is a generalization of classical logic, all classical relations defined in Hindi WordNet are incorporated in Fuzzy Hindi WordNet with a membership value of 1.

Formally, we define relations in Fuzzy Hindi WordNet as follows. Let W be a universal set of Hindi open class words/concepts. A relation $R \subseteq W \times W$ exists on a set W, where W includes all nouns, verbs, adjectives, and adverbs. Hindi WordNet can be represented as a graph with node set W and arc set R. Similarly, a fuzzy relation $\mu_R : W \times W \to [0, 1]$ can be represented as a fuzzy graph, where the arc $(x, y) \in W \times W$ has weight $\mu_R(x, y) \in [0, 1]$.

Formally, Fuzzy Hindi WordNet can be represented as a special fuzzy graph $G = (\sigma, \mu)$, which is a pair of functions $\sigma : W \to \{1\}$ and $\mu : W \times W \to [0, 1]$, where σ is the set of nodes that represents the set of Hindi open-class words (set of synsets) in which the strength of every node is 1. It means that every Hindi word fully belongs to Fuzzy Hindi WordNet. The μ denotes the strength of an edge between two Hindi words. The mapping shows the strength of relations between any two Hindi words.

Each entry in Fuzzy Hindi WordNet consists of a synset, gloss (description of the concept/synset), and position in ontology (the hierarchical organization of concepts), similar to Hindi WordNet. Now we propose the definition of the fuzzy relations in Fuzzy Hindi WordNet.

4.1. Relations in Fuzzy Hindi WordNet

Fuzzy Hindi WordNet is a word sense network. A word sense node in this network is a synset that is regarded as a basic object in Fuzzy Hindi WordNet. Each synset in Fuzzy Hindi WordNet is linked to other synsets through well-known lexical and semantic relations such as fuzzy hypernymy, fuzzy hyponymy, fuzzy meronymy, fuzzy troponymy, fuzzy antonymy, and fuzzy entailment. Semantic relations are between synsets, and lexical relations are between words. These relations serve to organize the lexical knowledge base.

There are 17 relations in Fuzzy Hindi WordNet. These relations are described next.

4.1.1. Relations between Same Parts of Speech.

Fuzzy Association

Hindi WordNet, which is a classical lexicon, provides synsets where each element of a synset has exactly similar meaning. In the Hindi language, we use in our day-to-day conversation many words that do not have exactly similar meaning but may have approximate similar meanings and can replace each other in many contexts, e.g., विद्यालय(vidyālaya) and संस्थान(saṁsthāna). In Fuzzy Hindi WordNet, if two concepts have approximately similar meanings, then they are related to fuzzy association relation. A fuzzy association relation is a semantic relation between two synsets to represent the partially similar meanings of two concepts. It can be written as

(विद्यालय, पाठशाला, स्कूल, पाठालय(vidyālaya, pāṭhaśālā, skūla, pāṭhālaya)) $\overset{0.8}{\to}$ (संस्थान, संस्था, अधिष्ठान, प्रतिष्ठान, इंस्टिट्यूट(saṁsthā, adhiṣṭhāna, pratiṣṭhāna, ïsṭiṭyūṭa)), where 0.8 represents the strength of the relation.

Formally, fuzzy association, represented as (w_1, w_2, μ_{as}), is defined as a fuzzy relation between two synsets (w_1 and w_2), where $w_1, w_2 \in W$ and μ_{as}: $W \times W \to [0, 1]$ such that w_1 and w_2 are approximate synonyms. Here, W denotes the set of all concepts/synsets and μ_{as} is the strength of fuzzy association relation.

Fuzzy association is a fuzzy relation that is reflexive, symmetric, and max-*-transitive.

Some more examples of fuzzy association are (अलमारी(alamārī), रैक(raika),0.7), (लडकी(laḍakī), औरत(aurata), 0.8), (लडका(laḍakā), आदमी(ādmī), 0.8), (चूडी(cūḍī), कंगन(kangana), 0.7), (मेज(mēja), स्टूल(sṭūla), 0.7), (किताब(kitāba), कुंजी(kuñjī), 0.7), (विद्यालय(vidyālaya), संस्थान(saṁsthāna), 0.8), (दस्ता(dastā), कॉपी(kêāpī), 0.7), (जीप(jīpa), कार(kāra), 0.7), (स्कूटर(skūṭara), मोटरसाईकिल(mōṭarasāīkila), 0.7), and (अटैची(aṭaicī), संदूक(sandūka), 0.7). Each word in these relations represents its synset; for the sake of brevity, we have used only one word from the synset.

Fuzzy Hypernymy and Fuzzy Hyponymy

Two synsets are said to share a fuzzy hypernymy/fuzzy hyponymy relationship if they approximately capture the superset/subset hood relationship. If $w_1, w_2 \in W$ are two

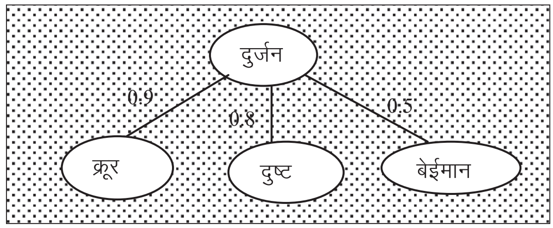


Fig. 1. The fuzzy hypernymy and fuzzy hyponymy relation.

Hindi synsets, then $(w_1, w_2, \mu_{hr}/\mu_{hp})$ is a fuzzy hypernym/hyponym relation between w_1 and w_2 if w_1 is an approximate subset of w_2 with a degree μ_{hr}/μ_{hp}, where $\mu_{hr}/\mu_{hp} : W \times W \rightarrow [0, 1]$.

Fuzzy hypernymy/fuzzy hyponymy is a semantic relation between two synsets to capture partial superset/subset hood, respectively. For example, there is a fuzzy hypernymy/hyponymy relation between मट्ठा(maṭṭhā) and दही(dahī), and it can be written as (मट्ठा, मठा, माठा, छाछ, तक्र, पादजल, अरिष्ट, मलिन, प्राग्राट (maṭṭhā, maṭhā, māṭhā, chācha, takra, pādajala, ariṣṭa, malina, prāgrāṭa)) $\xrightarrow{0.8}$ (दही, दधि, अमस्तु (dahī, dadhi, amastu)), where 0.8 is the strength of the fuzzy hypernymy/hyponymy relation.

Some more examples may include (दुखी(dukhī), परेशान(parēśāna), 0.9), (सज्जन(sajjana), दयालु(dayālu), 0.8), (पर्स(parsa), बैग(baiga), 0.6), (Teaching-cum-research-fellow,अध्यापक(adhyāpaka), 0.7), (क्रूर(krūra), दुर्जन(durjana), 0.9), (आईसक्रीम(āīskrīma), दूध(dūdha), 0.7), (आवाज(āvāja), ध्वनि(dhvani), 0.8), (पेपर.कटर(pēpara-kaṭara), हथियार(hathiyāra), 0.6), (सज्जन(sajjana), ईमानदार(īmānadāra), 0.8), (दुष्ट(duṣṭa), दुर्जन(durjana), 0.8), and (बेईमान(bēīmāna), दुर्जन(durjana), 0.5). Each word in these relations represents its synset; for the sake of brevity, we have used only one word from the synset.

Figure 1 shows the strength of the fuzzy hypernym/hyponym relationship between (क्रूर(krūra), दुर्जन(durjana)), (दुष्ट(duṣṭa), दुर्जन(durjana)), and (बेईमान(bēīmāna), दुर्जन(durjana)). The fuzzy hypernymy relation is the reverse of fuzzy hyponymy relation.

Fuzzy hypernymy ($\mu_{hr} : W \times W \rightarrow [0, 1]$) and fuzzy hyponymy ($\mu_{hp} : W \times W \rightarrow [0, 1]$) are fuzzy relations that are antireflexive, antisymmetric, and max-*-transitive.

Fuzzy Meronymy and Fuzzy Holonymy

Fuzzy meronymy and fuzzy holonymy are semantic relations that are used to capture the partial “part–whole” relation between two synsets, e.g., (गत्ता, कुट, दफ्ती, वसली (gattā, kuṭa, daftī, vasalī)) $\xrightarrow{0.7}$ (किताब, पुस्तक(kitāba, pustaka)). Every किताब(kitāba) does not have गत्ता(gattā), but many किताब(kitāba) do. Fuzzy meronymy is the reverse of fuzzy holonymy.

Fuzzy meronymy and fuzzy holonymy are the semantic relations represented as (w_1, w_2, μ_{me}), where $\mu_{me} : W \times W \rightarrow [0, 1]$ and $w_1, w_2 \in W$. Fuzzy holonymy (w_1, w_2, μ_{ho}) can be defined similarly.

Fuzzy meronymy and fuzzy holonymy are fuzzy relations that are antireflexive, antisymmetric, and max-*-transitive.

Some more examples may include (फल(phala), पेड, 0.7), (छात्रवास(chātravāsa), संस्थान(saṁsthāna), 0.8), (छात्रवास(chātravāsa), विद्यालय(vidyālaya), 0.6), (रेडियो(rēiḍayō), कार(kāra), 0.8), (कॉलर(kêālara),कमीज(kamīza), 0.8), (ऑले(ôlē), बारिश(bāriśa), 0.5), and (ज्वारभाटा(jvārabhāṭā), समुन्द्र(samundṛ), 0.5). Each word in these relations represents its synset.

Fuzzy Antonymy

Fuzzy antonymy is a lexical relation that holds between two words (not between two synsets) expressing approximately opposite meaning. Fuzzy antonymy is a fuzzy relation represented as (w_1, w_2, μ_{an}), where $\mu_{an} : W \times W \rightarrow [0, 1]$ and $w_1, w_2 \in W$. For example,

(बेईमान(bēīmāna)) $\xrightarrow{0.7}$ (सज्जन(sajjana)) where the strength of the relation is 0.7.

Some more examples may include (परेशान(parēśāna), खुश(khuśa), 0.8), (गुनगना(gunagunā), ठंडा(ṭhaṇḍā), 0.6), (गुनगना(gunagunā), गर्म(garma), 0.6), (हँसना(hêsanā), कराहना(karāhanā), 0.8), (बडा(badā), थोडा(thōḍā), 0.7), (ज्यादा(jyādā), छोटा(chōṭā), 0.7), (धुंधला(dhundhalā), चमकदार(camakadāra), 0.6), (क्रूर(krūra), सज्जन(sajjana), 0.7), and (दयालु(dayālu), बेईमान(bēīmāna), 0.7).

Fuzzy antonymy is a fuzzy relation $\mu_{an} : W \times W \rightarrow [0, 1]$ that is antireflexive, symmetric, and max-*-nontransitive.

Fuzzy Entailment

A verb A entails a verb B if the truth of B follows logically from the truth of A. It is a one-way relation. Fuzzy entailment (a semantic relation) holds between two synsets of verbs if the truth of verb B partially follows the truth of verb A.

Fuzzy entailment can be represented as $(v_1 \xrightarrow{\mu_e} v_2)$, where $\mu_e : V \times V \rightarrow [0, 1]$ and $v_1, v_2 \in V$, and $V (V \subset W)$ is the set of all verbs defined in the Hindi language. The μ_e is the strength/membership value of the fuzzy entailment relation. For example,

(सफलता मिलना, कामयाबी मिलना, विजय मिलना (saphalatā milanā, qāmayābī milanā, kāmayābī milanā, vijaya milanā)) $\xrightarrow{0.8}$ (परिश्रम करना, मेहनत करना, श्रम करना, उद्यम करना, मशक्कत करना, पसीना बहाना (pariśrama karanā, mēhanata karanā, śrama karanā, udyama karanā, maśakkata karanā, pasīnā bahānā)).

This means that acquiring "सफलता(saphalatā)" follows logically from doing "परिश्रम(pariśrama)" with a membership degree of 0.8.

Some more examples may be सोना(sōnā) $\xrightarrow{0.9}$ लेटना(lēṭanā), बदबू आना(badabū ānā) $\xrightarrow{0.7}$ गंदगी होना(gandagī hōnā), अच्छा व्यवहार होना(acchā vyavahāra hōnā) $\xrightarrow{0.8}$ अच्छे संसकार होना(acchē saṁskāra hōnā), स्वस्थ होना(svastha hōnā) $\xrightarrow{0.8}$ व्यायाम करना(vyāyāma karanā), चमकना(camakanā) $\xrightarrow{0.8}$ मेहनत करना(mēhanata karanā), and मेहनत करना(mēhanata karanā) $\xrightarrow{0.8}$ कष्ट सहना(kaṣṭa sahanā). Each word in these relations represents its synset.

Fuzzy entailment is a fuzzy relation $\mu_e : V \times V \rightarrow [0, 1]$ that is reflexive, antisymmetric, and max-*-transitive.

Fuzzy Troponymy

Fuzzy troponymy (a semantic relation) is a relation between the synset of two verbs, one denoting elaboration of another verb in an approximately specific manner.

The fuzzy troponymy relation can be represented as (v_1, v_2, μ_t), where $\mu_t : V \times V \rightarrow [0, 1]$ and $v_1, v_2 \in V$, and $V (V \subset W)$ is the set of all verbs defined in Hindi language. The μ_t is the strength/membership value of the fuzzy troponymy relation. For example,

(पढ़ना, पढ़ाई करना(paḍhanā, paṛhāī karanā)) $\xrightarrow{0.8}$ (सीखना(sīkhanā)).

This means that "पढ़ना, पढ़ाई करना(paḍhanā, paṛhāī karanā)" is an elaboration of "सीखना(sīkhanā)" with a membership degree of 0.8. Some more examples may be (दौडना(dauḍanā), चलना(calanā), 0.7) and (दुखी होना (dukhī hōnā), परेशान होना (parēśāna hōnā), 0.7). Each word in these relations represents its synset.

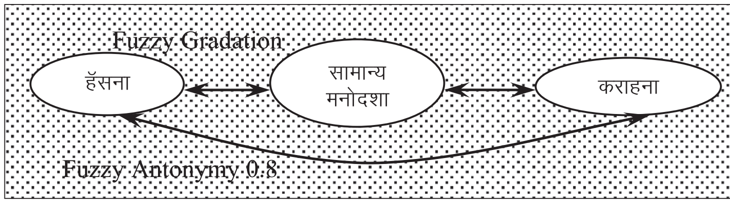


Fig. 2. Fuzzy gradation relation.

Fuzzy troponymy is a fuzzy relation $\mu_t : V \times V \rightarrow [0,1]$, which is antireflexive, antisymmetric, and max-*-transitive.

Note that fuzzy troponymy exists between two verbs, whereas fuzzy hypernymy/hyponymy exists between two nouns.

Fuzzy Gradation

Fuzzy gradation (a lexical relation) represents the intermediate concepts between fuzzy antonyms. For example, (हँसना(hêsanā), सामान्य मनोदशा(sāmānya manōdaśā), कराहना(karāhanā)) (Figure 2).

Fuzzy gradation is a fuzzy relation $\mu_g : W \times W \rightarrow [0,1]$ that is antireflexive, symmetric, and max-*-nontransitive.

Fuzzy Causative

The causative relation links the causative verbs and shows interdependency between them. It is a lexical relation, which means that it is a relation between two words and not between two synsets. The strength of a fuzzy causative relation is always unity, because this represents the relation between different morphological forms of a verb. For example, (खाना(khānā), खिलाना(khilānā), 1.0), खिलाना(khilānā) is a causative form of the verb खाना(khānā).

Fuzzy causative is a fuzzy relation $\mu_c : W \times W \rightarrow [0,1]$ that is reflexive, antisymmetric, and max-*-nontransitive.

Additionally, it can be noted that the following inferences hold on the previously defined relations:

(1) $(\mu_{as}(a,b) \neq 0) \rightarrow (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_c(a,b) = 0) \cap (\mu_{as}(b,a) = \mu_{as}(a,b))$

(2) $(\mu_{hr}(a,b) \neq 0) \rightarrow (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_t(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap (\mu_c(a,b) = 0) \cap (\mu_{hp}(b,a) = \mu_{hr}(a,b))$

(3) $(\mu_{hp}(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_t(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap \mu_c(a,b) = 0) \cap (\mu_{hr}(b,a) = \mu_{hp}(a,b))$

(4) $(\mu_{me}(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_e(a,b) = 0) \cap (\mu_t(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap (\mu_c(a,b) = 0) \cap (\mu_{ho}(b,a) = \mu_{me}(a,b))$

(5) $(\mu_{ho}(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_e(a,b) = 0) \cap (\mu_t(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap (\mu_c(a,b) = 0) \cap (\mu_{me}(b,a) = \mu_{ho}(a,b))$

(6) $(\mu_{an}(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_e(a,b) = 0) \cap (\mu_t(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap (\mu_c(a,b) = 0) \cap (\mu_{an}(b,a) = \mu_{an}(a,b))$

(7) $(\mu_e(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap (\mu_c(a,b) = 0)$

(8) $(\mu_t(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_g(a,b) = 0) \cap (\mu_c(a,b) = 0)$

(9) $(\mu_g(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_e(a,b) = 0) \cap (\mu_t(a,b) = 0) \cap (\mu_c(a,b) = 0) \cap (\mu_g(b,a) = \mu_g(a,b))$

(10) $(\mu_c(a,b) \neq 0) \rightarrow (\mu_{hr}(a,b) = 0) \cap (\mu_{hp}(a,b) = 0) \cap (\mu_{me}(a,b) = 0) \cap (\mu_{ho}(a,b) = 0) \cap (\mu_e(a,b) = 0) \cap (\mu_t(a,b) = 0)$

Let us explain inference (1) by considering an instance of association relation "(मेज(mēja), स्टूल(sṭūla),0.7)," i.e. ($\mu_{as}(b, a) \neq 0$ where a = मेज(mēja) and b = स्टूल(sṭūla), which means, मेज(mēja) shares association relation with स्टूल(sṭūla) with a degree of 0.7 (nonzero). From inference (1), we can infer that $\mu_{me}(a, b) = 0$ expresses स्टूल(sṭūla) can't be a part of मेज(mēja), मेज(mēja) can't be a part of स्टूल(sṭūla) is expressed by ($\mu_{ho}(a, b) = 0$), and there is no causative relation between मेज(mēja) and स्टूल(sṭūla) is expressed by ($\mu_c(a, b) = 0$) and स्टूल(sṭūla) also shares association relation with मेज(mēja) with an equal degree of 0.7, i.e., ($\mu_{as}(b, a) = \mu_{as}(a, b)$). Rest of the inferences can be explained in a similar way.

4.1.2. Relations between Cross Parts of Speech. In Fuzzy Hindi WordNet, fuzzy relations also exist between the synsets of different parts of speech. The following sections discuss various fuzzy relations that exist among different parts of speech.

Linkage between Nominal and Verbal Concepts

Fuzzy Ability Link

In Hindi WordNet, the ability link (a semantic relation) is used to specify the inherited features of a nominal concept. It would be better if a weight $\mu_{ab} \in [0, 1]$ is assigned to the ability link. For example, the ability "तैरना(tairanā)" of "मछली(machalī)" is more than the ability "तैरना(tairanā)" of "पेंगुइन(pēnguina)."

Fuzzy ability link (a semantic relation) can be represented as (n, v, μ_{ab}), where $n \in N$ and $N \subset W$ is the set of nouns, $v \in V$. where $V \subset W$ represents the set of verbs in the Hindi language and μ_{ab} represents the weight between n and v. The examples showing the fuzzy ability links are (मगरमच्छ(magaramaccha), तैरना(tairanā), 0.7), (पेंगुइन(pēnguina), तैरना(tairanā), 0.6), and (मछली(machalī), तैरना(tairanā), 1).

Fuzzy Capability Link

This link specifies the acquired features of a nominal concept. Fuzzy capability link (a semantic relation) is assigned a strength value $\mu_{cp} \in [0, 1]$ that shows the strength of the capability to acquire a feature.

Fuzzy capability link can be represented as (n, v, μ_{cp}) where $n \in N$ and $N \subset W$ is the set of nouns in Hindi language and $v \in V$, where $V \subset W$ represents the set of verbs in Hindi language.

For instance, the capability of वाहनचलाना(vāhana calānā) of a "आदमी(ādmī)" is more than the capability of वाहन चलाना(vāhana calānā) of a "लडका(laḍakā)" because in most of the cases "आदमी(ādmī)" refer to an adult whereas "लडका(laḍakā)" refer to a juvenile. A juvenile is not allowed to have a driving licence, therefore he would be less capable to वाहन चलाना(vāhana calānā) than a "आदमी(ādmī)." It may be represented as (आदमी(ādmī), वाहन चलाना(vāhana calānā), 0.9) and (लडका(laḍakā), वाहन चलाना(vāhana calānā), 0.6).

Fuzzy Function Link

This link specifies the function of a nominal concept. It is not justifiable to treat all functional links with equal strength (unity) as defined in Hindi WordNet. It can be better understood by considering an instance defined in Hindi WordNet: "पढाना(paṛhānā) is a function of a अध्यापक(adhyāpaka)." Ideally, all teachers are believed to teach; however, in reality, many teachers engaged in administrative work (like the vice chancellor of a university) are not assigned any teaching load. Moreover, some teachers are assigned a higher teaching load than others, e.g., a professor is assigned a lighter teaching load than an assistant professor or associate professor. On the contrary, professors are assigned more administrative and research responsibilities (and a lighter teaching load) than assistant professors and associate professors. The weight of a fuzzy function link is assigned on the basis of the strength of the respective function (Figure 3). Each word in these relations represents its synset.

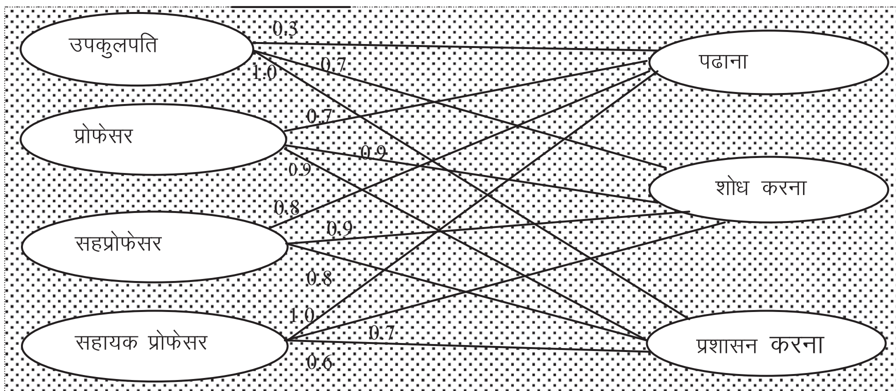


Fig. 3. Fuzzy function link.

Formally, a fuzzy function link (a semantic relation) can be represented as $(n, v\ \mu_f)$, where $n \in N$ and $N \subset W$ is the set of nouns in Hindi language, and $v \in V$, where $V \subset W$ represents the set of verbs in Hindi language.

Linkage between Nominal and Adjectival Concepts

Fuzzy Attribute

A fuzzy attribute denotes the partial properties of a noun. It is a semantic relation that represents the linkage between the synset of a noun and an adjective. For example, (लड़की, बालिका, बच्ची, बाला, कन्या, छोकरी, छोरी, पृथुका, टिमिली (laṛakī, bālikā, baccī, bālā, kanyā, chōkarī, chōrī, pṛthukā, ṭimilī)) $\xrightarrow{\mu at}$ (सुंदर, सुन्दर, खूबसूरत, ख़ूबसूरत, चारु, चारू, मनोहर, (sundara, sundara, k̲h̲ūbasūrata, khūbasūrata, cāru, cārū, manōhara)).

A fuzzy attribute can be represented as (n, a_d, μ_{at}), where $n \in N$ and $N \subset W$ is the set of nouns in the Hindi language, and $a_d \in A_d$, where $A_d \subset W$ is the set of adjectives in the Hindi language.

Some more examples are (जमीन(jamīna), उपजाऊ(upajāū), μ_{at}), (चाँदी(côdī), चमकदार (camakadāra), μ_{at}), (दिन(dina), साफ(sāpha), μ_{at}), and (पानी(pānī), गर्म(garma), μ_{at}), where the value of μ_{at} depends on the context. Each word in these relations represents its synset; again, for the sake of brevity, we have used only one word from the synset.

Fuzzy Modifies Noun

Some adjectives can only go with certain nouns. Such adjectives and nouns are linked in Fuzzy Hindi WordNet by the modifies noun relation. For example, (सुपात्र,सत्पात्र,अच्छा पात्र (supātra,satpātra,acchā pātra)) is the adjective that can only be used to describe a (व्यक्ति,मानस,शख्स,शख़्स,जन,बंदा,बन्दा (vyakti,mānasa,śak̲h̲sa,śakhsa,jana,bandā,bandā)). The membership value of this relation is always unity, i.e., it is a semantic relation between two synsets that always act as a classical relation.

It can be represented as $(n, a_{dj}, 1)$, where $n \in N$ and $N \subset W$ is the set of nouns in the Hindi language, and $a_{dj} \in A_d$, where $A_d \subset W$ is the set of adjectives in the Hindi language.

Linkage between Adverbial and Verbal Concepts

Fuzzy Modifies Verb

The fuzzy modifies verb relation (a semantic relation) shows the connection between adverb and verb such that the adverb modifies only this specific verb. For example, (तेज, तेज़, तेजी से, तेज़ी से, रफ्तार से, रफ़्तार से, तेज गति से (tēja, tēza, tēzī sē, tējī sē, raftāra sē, raphtāra sē, tēja gati sē)) $\xrightarrow{1.0}$, (दौड़ना, भागना, धाना (dauṛanā, bhāganā, dhānā)) तेज(tēja) is used to describe भागना(bhāganā), so the membership value of this relation is always unity, i.e., it is always a classical relation.

Fuzzy Derived From

The fuzzy derived from relation is a lexical relation. Fuzzy derived from specifies the root form from which a particular word is derived. For example, (क्रमशः(kṛmaśaḥ), क्रम(kṛma), 1) क्रमशः(kṛmaśaḥ) is the adverb that is used to describe the verb क्रम(kṛma) only. The membership value of this relation is always unity, i.e., it is always a classical relation.

All relations between cross parts of speech are the fuzzy relations μ: W $\times$ W $\rightarrow$ [0, 1], which are antireflexive, antisymmetric, and max-*-nontransitive.

We now discuss the method to obtain the membership values between any two concepts.

4.2. Assigning Membership Value between the Pair of Words for Relations in Fuzzy Hindi WordNet

In the literature, the method of expert opinion has been used to assign the membership value between any two concepts in a fuzzy relation. In this method, multiple experts in the domain are consulted to derive the membership value between two concepts. Chen et al. [2001] used expert opinion for document retrieval to assign the membership value to the fuzzy relations between concepts in fuzzy-valued concept networks. In 2003, researchers again extended their work for fuzzy information retrieval based on multirelationship fuzzy concept networks and used expert opinion to establish the membership value in multirelationship fuzzy concept networks [Chen et al. 2003]. Recently, Tayal et al. [2014] also used the expert opinion method to assign the fuzzy membership value to the academic expertise of a reviewer. The authors then proposed an approach to identify the best reviewer suited to evaluate a submitted research proposal. In the method of expert opinion, a membership value $\mu(C_i, C_j) \in [0, 1]$ is assigned between any two concepts C_i and C_j based on the opinion of the domain experts.

To assign the strength to the fuzzy relations between words proposed in Fuzzy Hindi WordNet, we involved selected renowned Hindi lexicographers—experts in linguistics, Hindi language, computer science, and fuzzy logic. Every expert at the meeting was required to rate a value ranging between 0 and 1 for a relation between every pair of words (w_i, w_j). As per the expert's own perception and understanding, this value represented the strength of the relation between words w_i and w_j. The final strength for the two words in various relations was calculated by taking the average of the values assigned by experts. The similar process was repeated for obtaining the membership values between all Hindi words in various fuzzy relations proposed in Fuzzy Hindi WordNet.

4.3. Composition of Relations in Fuzzy Hindi WordNet

One main feature of Fuzzy Hindi WordNet is that it is not necessary to explicitly provide relations between all pairs of concepts/words. In fact, here each word is connected to every other word (albeit fuzzily). The relation type and strength of connection between two words can be computed by using the concept of composition proposed in this section. By using the inherent transitivity in the relations, we can compute relations between semantically related words even if it is not explicitly given. Note that the strength of connection between any pair of words, say w_1 and w_2, may be "0," i.e., the words

Table I. Matrix M

X_i \ Y_j	FAS	FHR	FHP	FME	FHO	FAN	FEN	FTR
FAS	FAS	FHR	FHP	FME	FHO	FAN	FEN	FAS
FHR	FHR	FHR	FAS	FME	FHO	FAN	-	-
FHP	FHP	FAS	FHP	FME	FHO	FAN	-	-
FME	FME	FME	FME	FME	FAS	-	-	-
FHO	FHO	FHO	FHO	FAS	FHO	-	-	-
FAN	FAN	FAN	FAN	-	-	FAS /FAN	FAN	FAN
FEN	FEN	-	-	-	-	FAN	FEN	FEN
FTR	FAS	-	-	-	-	FAN	FEN	FTR

FAS	Fuzzy Association	**FHO**	Fuzzy Holonymy
FHR	Fuzzy Hypernymy	**FAN**	Fuzzy Antonymy
FHP	Fuzzy Hyponymy	**FEN**	Fuzzy Entailment
FME	Fuzzy Meronymy	**FTR**	Fuzzy Troponymy

Table II. Matrix S

X_i \ Y_j	FAS	FHR	FHP	FME	FHO	FAN	FEN	FTR
FAS	O	M	M	P	P	M	P	P
FHR	M	O	M	M	M	M	-	-
FHP	M	M	O	M	M	M	-	-
FME	P	M	M	O	P	-	-	-
FHO	P	M	M	P	O	-	-	-
FAN	M	P	P	-	-	M	P	P
FEN	P	-	-	-	-	P	O	P
FTR	P	-	-	-	-	P	P	O

P	Pessimistic
M	Moderate
O	Optimistic

w_1 and w_2 are not connected at all (as in the case of Hindi WordNet). Here we not only consider the directly linked concepts but also traverse one or more links to find more related concepts. However, for computing the semantic closeness of two concepts, we consider the strength and relation types of the traversed links.

Table I (matrix M) shows the results of combining relations of different types, and Table II (matrix S) shows the corresponding strength of the combined relation.

Table I can be viewed as an 8 × 8 matrix (M) holding the following inference:

$$\forall (a, b) \in X_i \;\&\&\; \forall (b, c) \in Y_j \rightarrow (a, c) \in M(i, j), \tag{1}$$

where $1 = i = 8$, $1 = j = 8$ and $M(i, j)$ is the element in matrix M at the i^{th} row and j^{th} column.

Here, a, b, and c represent the concepts, and X_i, Y_j, and $M(i, j)$ represent the relations between concepts. If there is an X_i relation between concepts a and b and a Y_j relation between concepts b and c, then there exists a relation $M(i, j)$ between concepts a and c. The type of relation $M(i, j)$ and its corresponding strength can be obtained from Tables I and, respectively. This can be further explained by considering the following example.

Example 1. (वाहन(vāhana), कार(kāra)) shares fuzzy hypernymy relation X_2, and (कार(kāra), रेडियो(rēiḍayō)) shares fuzzy meronymy relation Y_4. The relation type between वाहन(vāhana) and रेडियो(rēiḍayō) can be obtained from Table I and comes out to be fuzzy meronymy (*M*(*2,4*)). The strength of the corresponding relation can be obtained from Table II (*S*(*2,4*)) and comes out to be moderate. Therefore, it conveys that वाहन(vāhana) and रेडियो(rēiḍayō) are moderately related. In other words, it means that in a moderate number of cases, a "वाहन(vāhana)" has a "रेडियो(rēiḍayō)," i.e., a "रेडियो(rēiḍayō)" is a part of "वाहन(vāhana)" with a moderate strength (fuzzy meronymy). Note that in Hindi WordNet, no relation can be defined between वाहन(vāhana) and रेडियो(rēiḍayō). Therefore, it becomes difficult to deal with sentences like "वाहन चलाते समय रेडिया सुनने का आनन्द ही अलग है (vāhana calātē samaya rēiḍayō sunnē kā ānnada hī alaga hai)" using Hindi WordNet, but this type of sentence can easily be processed using Fuzzy Hindi WordNet.

After discussing the composition of fuzzy relations, we now propose the concepts for computing the precise strength of composed fuzzy relations using t-norms, which are monotone, commutative, and associative functions $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ that are used for fuzzy set intersection [Klir and Folger 1998]. Although a number of t-norm operators have been proposed in the literature, Bonissone and Keith [1986] have shown that for practical purposes, three to five different t-norm operators are sufficient to obtain the results of realistic precision. As a useful set of operators that are good representatives of the range of possible t-norm functions, Bonissone and Keith [1986] proposed the following t-norms:

$$\begin{aligned} &T_1(x, y) = max(0, x + y - 1)\\ &T_2(x, y) = xy\\ &T_3(x, y) = min(x, y) \end{aligned} \tag{2}$$

Suppose that two relations exist with their corresponding strengths as $x = 0.8$ and $y = 0.5$. Depending on the t-norm, we can compute the strength of the composed fuzzy relation as 0.3 using T_1, 0.4 using T_2, or 0.5 using T_3.

It can be easily seen that $T_1(x, y) = T_2(x, y) = T_3(x, y)$ holds. Therefore, it can be concluded that the operators T_1 makes quite pessimistic results, T_2 makes moderate results, and T_3 makes rather optimistic results. On applying these t-norms on the composition of fuzzy relations defined earlier, the following observations can be made:

(1) It is better to use an optimistic t-norm operator for composition of fuzzy relations of the same type, with exception of fuzzy antonymy.
(2) If a fuzzy association relation is composed with any other relation, it results in that relation only, with exception of fuzzy troponymy.
(3) When fuzzy hypernymy (or fuzzy hyponymy) is composed with any other relation, the resultant fuzzy relation always possesses moderate strength.

5. WORD SENSE DISAMBIGUATION USING FUZZY HINDI WORDNET

WordNet [Fellbaum 1998] is widely used for WSD [Navigli and Lapata 2010; Erk et al. 2013; Kotle and Bhairal 2009a, 2009b; Banerjee and Pedersen 2002; Mihalcea 2005; Sinha and Mihalcea 2007; Mihalcea and Dragmoir 2011]. In a similar way, Hindi WordNet increasingly is being used for WSD in the Hindi language [Jain et al. 2013; Mishra et al. 2009; Sinha et al. 2004; Avneet 2010; Sinha and Siddiqui 2012; Sandeep and Chanchal 2012]. WSD refers to identifying the meaning of words in a context computationally. Virtually all WSD methods heavily rely on knowledge—either corpora or dictionaries. Therefore, the so-called knowledge acquisition bottleneck is undoubtedly one of the most important issues in WSD [Navigli 2009].

In this section, we show that WSD can be achieved at a higher level by using Fuzzy Hindi WordNet and fuzzy graph connectivity measures. Fuzzy Hindi WordNet can be viewed as a fuzzy graph where nodes represent concepts/synsets and edges represent fuzzy relation between concepts/synsets. Relation between any two concepts may be any of the fuzzy relations defined in Fuzzy Hindi WordNet. Recently, Jain et al. [2013] proposed a method for WSD by using graph connectivity measures and Hindi WordNet. In Hindi WordNet, an edge represents the relation between concepts, and the weight of the edge (which is always unity) represents the distance between concepts. In Fuzzy Hindi WordNet, the weight of an edge represents the strength of the relation between concepts, and its value always lies in [0, 1]. Jain et al. [2013] used the local and global connectivity measures for Hindi WordNet.

For a fuzzy graph, no local and global connectivity measures have yet been defined in the literature, and thus we first hereby propose local and global connectivity measures for a fuzzy graph. Since Fuzzy Hindi WordNet can be rendered as a fuzzy graph, we will use these measures for disambiguation of Hindi language sentences.

5.1. Fuzzy Graph Connectivity Measures

We propose the local and global measures for a fuzzy graph as follows.

5.1.1. Local Measures for a Fuzzy Graph. Local measures of graph connectivity determine the degree of relevance of a node in the graph. Therefore, they can be viewed as a measure of the influence of vertex over the network. We define local measure (lm) as $lm: V \rightarrow [0, 1]$, where a value close to 0 indicates that the vertex or node is peripheral, and a value close to 1 indicates that it is important. For a graph, several local measures of graph connectivity based on centrality have been proposed in the literature [Freeman 1979; Borgatti 2006; Newman 2005; Opsahl et al. 2010; Freeman et al. 1991], where the importance of a node indicates that the node is central, i.e., maximally connected to all other nodes. However, for fuzzy graphs, no centrality measures have yet been defined in the literature.

A special fuzzy graph $G = (V, E)$ is a pair of functions $V : W \rightarrow \{1\}$ and $E : W \times W \rightarrow [0, 1]$, where V is the set of nodes (synsets) that represents the set of Hindi open-class words W (noun, verb, adjective, adverb) where the strength of every node is always 1. In this section, we propose four well-known measures of centrality, namely degree, eigenvector (PageRank and HITS), closeness, and betweenness centrality, for the fuzzy graph.

5.1.1.1. Degree Centrality. The degree of a node is the simplest centrality measure and is defined as the number of edges terminating at the given node. Zadeh et al. [1975] defined the degree of a vertex in a fuzzy graph (G) as follows. The degree of a vertex v (represented as $deg_f(v)$) in a fuzzy graph is defined as

$$deg_f(v) = \sum_{u \neq v} \mu_{uv} \; where \; (u, v) \in E, \text{the set of edges.} \tag{3}$$

We define the degree centrality ($C_f(v)$) of a vertex v as a ratio of the degree of a vertex to the possible maximum degreeof a vertex in the fuzzy graph:

$$C_f(v) = \frac{deg_f(v)}{|v| - 1}, \tag{4}$$

5.1.1.2. Eigenvector Centrality. Eigenvector centrality assigns relative scores to all nodes in a graph based on the recursive principle that the connections having high scores contribute more to the score of the node to which it is connecting. PageRank and HITS are the variants of the eigenvector centrality measure and have been widely used in graph-based WSD [Jain et al. 2013]. In the present work, we first propose PageRank and HITS measures for fuzzy graphs and then use them for WSD using Fuzzy Hindi WordNet, which is represented as a fuzzy graph.

For fuzzy graphs, given a weight μ_{ab} of an edge connecting vertices v_a and v_b, we propose the definition for s PageRank score for the vertex v_a as

$$PageRank_f(v_a) = \frac{(1-d)}{|V|} + d \sum_{(v_a, v_b) \in E} \frac{\mu_{ba}}{\sum_{(v_b, v_c) \in E} \mu_{bc}} PageRank_f(v_b), \tag{5}$$

where E is the set of edges in the fuzzy graph, μ_{ij} represents the edge strength between vertex v_i and v_j, and $d \in [0, 1]$ is the damping factor. Luca and Carlos [2006] advocate that taking the value of d around 0.85 gives better results, and therefore we have also taken $d = 0.85$ in this article. The initial PageRank score of each vertex is considered to be 1.

HITS is another graph-centrality algorithm that was designed for ranking Web pages according to their degree of "authority" [Kleinberg 1999, Mihalcea and Dragmoir 2011].

HITS is similar to the PageRank algorithm; however, the main difference is that it makes a distinction between authorities and hubs. HITS [Kleinberg 1999] determines two values for each node v, the authority value $a(v)$, and the hub value $h(v)$. For fuzzy graphs, we propose the following definition for the hub and the authority values as follows:

$$h_f(v) = \sum_{(u,v)\in E} \mu_{uv} a_f(u) \tag{6}$$

$$a_f(v) = \sum_{(u,v)\in E} \mu_{uv} h_f(u), \tag{7}$$

where μ_{uv} is the strength of an edge between node u and v. Normalized values of hub $h'_f(v)$ and authority $a'_f(v)$ of a vertex can be obtained by dividing the hub value (authority value) by the sum of the square root of the hub (authority) values of all nodes as given next [Gupta 2006]:

$$h'_f(v) = \frac{h_f(v)}{\sqrt{\sum_{\forall i\in V} h_f(i)^2}} \tag{8}$$

$$a'_f(v) = \frac{a_f(v)}{\sqrt{\sum_{\forall i\in V} a_f(i)^2}}. \tag{9}$$

5.1.1.3. Closeness Centrality. For a classical graph, closeness centrality states that a vertex is important if it is relatively close to all other vertices [Borgatti 2006; Mihalcea and Dragmoir 2011]:

$$\text{closeness } centrality(u_n) = \frac{\sum_{\mathrm{u_0}\in V,\mathrm{u_0}\neq \mathrm{u_n}} \frac{1}{d(\mathrm{u_0},\mathrm{u_n})}}{|v|-1}, \tag{10}$$

where, $d(u_0, u_n)$ is the shortest distance between vertex u_0 and vertex u_n. The numerator in the equation is the sum of the inverse of the shortest distances between u_n and all other vertices, and the denominator is the number of nodes in the graph (excluding u_n).

For fuzzy graphs, the path length $\delta(u_0, u_n)$ of a path $P : u_0 u_1 \ldots\ldots u_n$ between any two nodes u_0 and u_n is calculated as follows [Rosenfeld et al. 1975; Sunitha 2001]:

$$\delta(u_0, u_n) = \sum_{i=1}^{n} \frac{1}{\mu_{(i-1)(i)}} \tag{11}$$

if $n = 0$, then $\delta(u_0, u_n) = 0$.

The shortest distance between u_0 and u_n in a fuzzy graph would be $d(u_0, u_n)$ and defined as

$$d(u_0, u_n) = \underset{all\ paths\ u_0\ to\ u_n}{Min} [\delta(u_0, u_n)]. \tag{12}$$

We propose the following definition for the closeness centrality of a vertex u_n for a fuzzy graph as follows:

$$closenesscent_f(u_n) = \frac{\sum_{u_0\in V, u_0\neq u_n} \frac{1}{\underset{all\ paths\ u_0\ to\ u_n}{Min} [\delta(u_0,u_n)]}}{|v|-1}. \tag{13}$$

The closeness centrality of a disconnected node is a small constant given by $\frac{1}{|V|}$.

5.1.1.4. Betweenness Centrality. The intuition behind betweenness is that a node is important if it falls in a large number of paths compared to the total set of paths. Betweenness of a vertex v is calculated as the fraction of shortest paths between node pairs that pass through v [Freeman 1979; Freeman et al. 1991]:

$$betweenness_f(v) = \sum_{s,t \in V, s \neq v \neq t} \frac{|d_v(s,t)|}{|d(s,t)|}, \tag{14}$$

where $|d(s,t)|$ represents the total number of shortest paths between s and t, and $|d_v(s,t)|$ represents the number of shortest paths from s to t that pass through vertex v. For a vertex v in a fuzzy graph, *betweenness centrality*(v) can be defined as the ratio of *betweenness*(v) to maximum number of node pairs excluding v.

$$betweennesscent_f(v) = \frac{betweenness_f(v)}{(|V|-1)(|V|-2)} \tag{15}$$

5.1.2. Global Measures for a Fuzzy Graph. Global connectivity measures are concerned with the structure of a graph as a whole rather than with individual nodes [Jain et al. 2013]. They are calculated on subgraphs derived from the main graph. We now propose the global connectivity measures for a fuzzy graph.

5.1.2.1. Compactness. Compactness refers to the extent of cross referencing in a graph. When compactness is high, reachability between vertices is good. A low compactness shows insufficient links and also indicates that parts of the network are possibly disconnected. For a graph G, compactness $C(G)$ is given by Jain et al. [2013] and Botafogo et al. [1992]:

$$C(G) = \frac{max - \sum_{u \in V} \sum_{v \in V} \rho(u,v)}{max - min}. \tag{16}$$

In the equation, $\rho(u,v)$ is the shortest distance between node u and node v. The value of $max = max(\sum_{v \in V} \sum_{v \in V} \rho(u,v))$ is the maximum value the distance sum can assume, i.e., a completely disconnected graph. This distance sum can be approximated as $max = K|V||V-1|$, where $K = |V|$ for a graph. The value of $min = |V||V-1|$, the minimum value, i.e., a fully connected graph.

In the fuzzy graph, the distance between two neighbor nodes is taken as the inverse of the edge strength $\in [0,1]$, i.e., $e = 1/$(edge strength). Assuming that the *min* value of the strength of any of the relations defined in Fuzzy Hindi WordNet is 0.5, the maximum distance between two neighbor nodes can be computed as $1/(min$ edge strength$) = 1/0.5 = 2$. So the value of $max = 2|V||V||V-1|$. It may be noted that the value of *min* for a fuzzy graph remains similar to the graph, i.e., $min = |V||V-1|$. Now we propose the following definition for compactness $C_f(G)$ of a fuzzy graph as follows:

$$C_f(G) = \frac{2|\mathrm{V}|^2|\mathrm{V}-1| - \sum_{u \in V} \sum_{v \in V} d(u,v)}{2|\mathrm{V}|^2|\mathrm{V}-1| - |\mathrm{V}||\mathrm{V}-1|}, \tag{17}$$

where the value of $d(u,v)$ can be computed using Equation (12).

5.1.2.2. Graph Entropy. Entropy measures the amount of information (uncertainty) in a random variable. In graph-theoretic terms, high entropy indicates that many vertices are equally important, whereas low entropy indicates that only a few vertices are important. Entropy in a graph is given by Alon and Orlitsky [1996]:

$$H(G) = -\sum_{v \in V} p(v) log(p(v)), \tag{18}$$

where probability $p(v)$ of a vertex v is given by the degree distribution:

$$p(v) = \frac{\deg(v)}{2|E|} \tag{19}$$

We know that for fuzzy graphs, the degree of a node can be calculated as $\deg_f(v) = \sum_{u \neq v} \mu_{uv}$, where $(u, v) \in E$.

We define total number of edges ($|E|_f$) in a fuzzy graph:

$$|E|_f = \sum_{u \in V} \sum_{v \in V} \mu_{uv}, \tag{20}$$

where $\mu_{uv} = 0\, if\, (u, v) \notin E$.

We propose the definition for graph entropy of a fuzzy graphs $H_f(G)$ as follows:

$$H_f(G) = -\sum_{v \in V} \frac{\sum_{u \neq v} \mu_{uv}}{2|E|} log \frac{\sum_{u \neq v} \mu_{uv}}{2|E|}, \tag{21}$$

where $(u, v) \in E$, $H_f(G)$ provides a measure of importance of nodes in a fuzzy graph. However, $H_f(G)$ is not normalized. Normalized entropy for a fuzzy graph can be obtained by dividing $H_f(G)$ by the maximum entropy given by $log|V|$.

5.1.2.3. Edge Density. In a graph G with $|V|$ vertices, edge density is defined as the ratio of edges in a graph to the number of edges of a complete graph [Jain et al. 2013]:

$$ED(G) = \frac{|E(G)|}{|V|_{C_2}}. \tag{22}$$

For fuzzy graphs, the number of edges is calculated by using Equation (20), i.e., by summing up the all edge strengths. So we propose the definition of edge density for a fuzzy graph G as

$$ED_f(G) = \frac{\sum_{u \in V} \sum_{v \in V} \mu_{uv}}{|V|_{C_2}}. \tag{23}$$

5.2. Word Sense Disambiguation Using Fuzzy Graph Connectivity Measures

In this section, we apply the concepts of Fuzzy Hindi WordNet and fuzzy graph connectivity measures proposed by us in the previous sections for WSD in the Hindi language. Recently, Jain et al. [2013] proposed a WSD method for all open-class words in the Hindi language. In this article, we hereby propose a modified WSD algorithm (Algorithm 1) using Fuzzy Hindi WordNet and fuzzy graph connectivity measures.

Initially, we build a fuzzy graph $G = (V, E)$, named as sentence graph for each target sentence, by using the graph of the reference lexicon Fuzzy Hindi WordNet. Sentences are part-of-speech tagged (POS tagged), as our algorithm considers content words only, i.e., nouns, verbs, adjectives, and adverbs. As explained earlier, the vertices in a graph are word senses, and the edges represent semantic relations. Given the fuzzy graph G, for each content word $w_i \in W$, W represents the set of open-class words extracted from the given sentence. We select the most appropriate sense $S_{w_i} \in Senses(w_i)$, where $Senses(w_i)$ is the set of senses of w_i listed in reference lexicon (Fuzzy Hindi WordNet). We identify appropriate sense $S_{w_i} \in Senses(w_i)$ by ranking each vertex in the graph G according to its importance, which is computed using fuzzy graph connectivity measures.

The following example demonstrates how the proposed WSD algorithm can be used for disambiguating word senses in a sentence in the Hindi language.

ALGORITHM 1: WSD Using Fuzzy Hindi WordNet

Input: W (a set of open-class POS-tagged Hindi words extracted from the given Hindi sentence)
Let $W = (w_1, w_2, \ldots . w_i \ldots w_n)$, where w_i represents the $i^{th}((1 \le i \le n))$ word in the input, where $n = |W|$.
Step 1: Construct the sentence graph $G = (V, E)$ for the word sequence W using Fuzzy Hindi WordNet by using the following the steps:

i. Find all senses associated with each word $w_i \in W$ from Fuzzy Hindi WordNet. Let $senses(w_i)$ denote the set of all senses for w_i, and let V_W denote the collection of these senses for all $w_i \in W$, i.e.,

$$V_W = \bigcup_{i=1}^{n} senses(w_i)$$

ii. Initialize sentence graph $G(V, E)$:

$$V := V_W \text{ and } E := \emptyset.$$

iii. Taking each node $v \in V_W$ as a source node, repeat the following steps:
   a. Perform depth first search (DFS) of the Fuzzy Hindi WordNet graph considering v as source node up to length $\geq L$.
   Note: A moderate value for L should be chosen (usually $L \in [5, 8]$), because if a lower integer value of L is taken, then the required path between words of the given sentence is not established. However, if a higher value of L is taken. then the context is dissolved.
   b. During the traversal, if a node v' is encountered such that $v' \in V_W$, $v' \neq v$, and v and v' are not polysemous, then add the path from v to v' $(v, v_1, v_2, \ldots, v_k, v')$ in G, i.e., add all nodes and edges that are encountered in the path:
   $V = V \bigcup \{v_1, v_2, \ldots v_k\}$
   $E = E \bigcup \{(v, v_1), (v_1, v_2), \ldots (v_k, v')\}$;

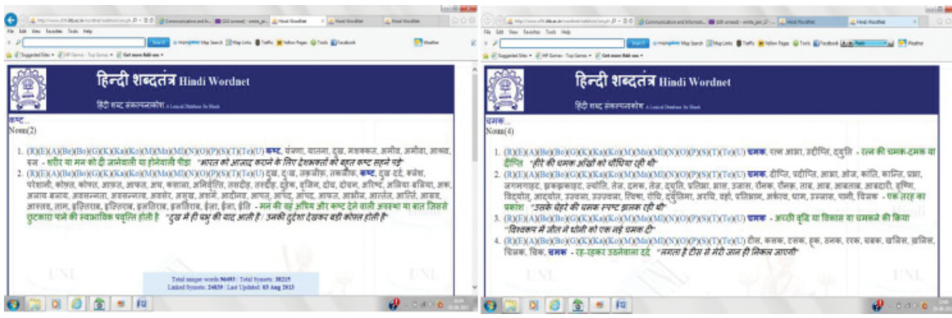
Step 2: For the sentence graph $G = (V, E)$ obtained in step 1, Compute all local connectivity measures (various centrality measures) for each node (using the equations proposed in Section 5.1.1).
Step 3: For each interpretation of the sentence, construct an interpretation graph and also compute global connectivity measures for each interpretation graph by using the following the steps:

i. Let each $w_i \in W(1 \le i \le n)$ have m_i senses defined in Fuzzy Hindi WordNet. The given sentence has $I = \prod_{i=1}^{n} m_i$ interpretations. For each interpretation, construct the interpretation graph $G_j(V_j, E_j) \subseteq G(V, E)(1 \le j \le I)$ by following these steps:
   a. Initialize $G_j\ (V_j, E_j)$:
   $V_j :=$ all nodes (word senses) representing the interpretation, $E_j := \emptyset$.
   b. For every node pair $r, s \in V_j$, let V_{rs}, E_{rs} be the set of nodes and set of edges respectively encountered in the path connecting node r and node s in G:
   Let $V'_j = \bigcup_{\forall (r,s) \in V_j} V_{rs}$, $\quad E'_j = \bigcup_{\forall (r,s) \in V_j} E_{rs}$
   $V_j = V_j \bigcup V'_j$, $\quad E_j = E_j \bigcup E'_j$

ii. Compute all global connectivity measures for each interpretation graph $G_j(V_j, E_j)(1 \le j \le I)$ (using the equations proposed in Section 5.1.2).

Step 4: For each content word $w_i \in W$, rank the word senses (vertices in G) according to its importance, which is computed using fuzzy graph connectivity measures (i.e., local and global) obtained in step 2 and step 3. Select the most appropriate word sense $S_{w_i} \in Senses\,(w_i)\ (1 \le i \le n)$ having the highest rank among the vertices.

Example 2. Consider the Hindi sentence "कष्ट झेलकर ही चमकते है। (kaṣṭa jhēlakara hī camakte hai.)." The content words are "कष्ट(*kaṣṭa*)," "झेलकर(*jhēlakara*)," and "चमकते(camakte)." Therefore, W = (कष्ट$_n$($kaṣṭa_n$), झेलना$_v$($jh\bar{e}lan\bar{a}_v$), चमक$_n$($camak_n$)), and $|W| = 3$.



(a) (b)

Fig. 4. Meanings of the words "कष्ट" and "चमक" in Hindi WordNet.

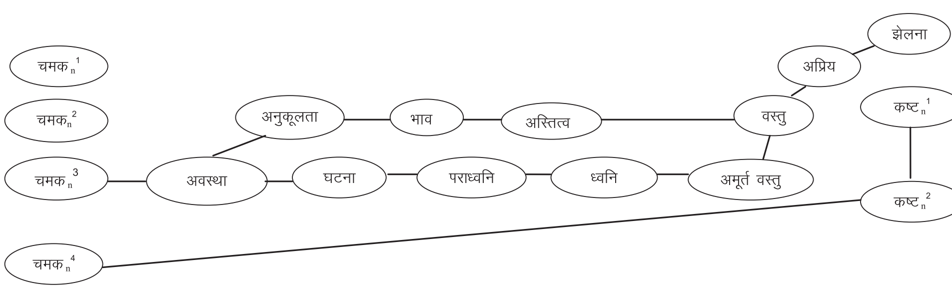


Fig. 5. Sentence graph constructed for the sentence "*कष्ट झेलकर ही चमकते है*" using Hindi WordNet.

For the words "*कष्ट*(*kaṣṭa*)," "*झेलकर*(*jhēlakara*)," and "*चमकते*(camakte)," Fuzzy Hindi WordNet and Hindi WordNet list 2, 1, and 4 senses, respectively (here, the subscript n represents a noun and the subscript v represents a verb). First we extract the base words from the content words, i.e., "*झेलना*(*jhēlanā*)" for "*झेलकर*(*jhēlakara*)" and "*चमक*(*camak*)" for "*चमकते*(camakte)," the word "*कष्ट*" remains similar. Refer to Figure 4 for a description of different meanings of the words "*कष्ट*(*kaṣṭa*)" and "*चमक*(*camak*)" taken from Hindi WordNet. Initially, we set

$V_W = \{$कष्ट$_n^1$(*kaṣṭa*$_n^1$), कष्ट$_n^2$(*kaṣṭa*$_n^2$), झेलना$_v^1$(*jhēlanā*$_v^1$), चमक$_n^1$(*camak*$_n^1$), चमक$_n^2$(*camak*$_n^2$), चमक$_n^3$(*camak*$_n^3$), चमक$_n^4$(*camak*$_n^4$)$\}$, initializing $G(V,E) V := V_w$ and $E = \emptyset$. Now we perform DFS taking each node in V_W as a source node one by one. We take $L = 8$. The node कष्ट$_n^1$(*kaṣṭa*$_n^1$) is adjacent to कष्ट$_n^2$(*kaṣṭa*$_n^2$) in Hindi WordNet by connecting through the hypernymy/hyponymy relation. We first follow कष्ट$_n^1$(*kaṣṭa*$_n^1$), which is in turn connected to कष्ट$_n^2$(*kaṣṭa*$_n^2$); later, it is connected to चमक$_n^4$(*camak*$_n^4$). We add the path (कष्ट$_n^1$(*kaṣṭa*$_n^1$), कष्ट$_n^2$(*kaṣṭa*$_n^2$), चमक$_n^4$(*camak*$_n^4$)) in graph G. Similarly, when we traverse the Hindi WordNet graph taking चमक$_n^3$(*camak*$_n^3$) as a source node, which is connected with झेलना$_v^1$(*jhēlanā*$_v^1$) by two different paths {चमक$_n^3$(camak$_n^3$), अवस्था(avasthā), अनुकूलता(anukūlatā), भाव(bhāva), अस्तित्व(astitva), वस्तु(vastu), अप्रिय(apiṛya), झेलना(jhēlanā)} and {चमक$_n^3$(camak$_n^3$), अवस्था(avasthā), घटना(ghaṭan), पराध्वनि(parādhvani), ध्वनि(dhvani), अमूर्तवस्तु(amūrtavastu), वस्तु(vastu), अप्रिय(apiṛya), झेलना(jhēlanā)} of path lengths 7 and 8, respectively. As चमक$_n^3$(*camak* $_n^3$), झेलना$_v^1$(*jhēlanā*$_v^1$) ϵ V_W, we add all intermediate nodes and edges encountered in these paths in the sentence graph. Following a similar procedure, we first construct the sentence graph (Figure 5) using Hindi WordNet and simultaneously construct the sentence graph (fuzzy graph) (Figure 6) using Fuzzy Hindi WordNet. Comparing these two figures, we notice that the fuzzy graph in Figure 6 shows

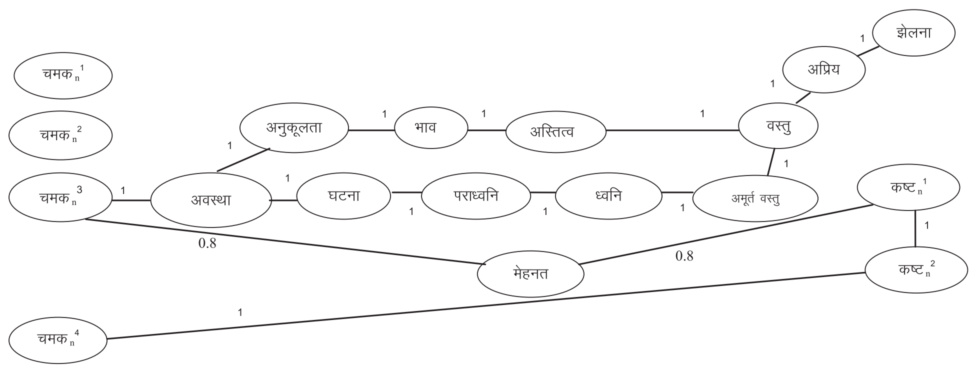


Fig. 6. Sentence graph constructed for the sentence *“कष्ट झेलकर ही चमकते है”* using Fuzzy Hindi WordNet.

Table III. Values of Local Connectivity Measures for the Sentence Graph Shown in Figure 5 (Using Hindi WordNet)

Words	Degree	PageRank	HITS	Closeness	Betweenness
चमक$_n^1$	0.0000	0.0088	0.0000	0.0625	0.0000
चमक$_n^2$	0.0000	0.0088	0.0000	0.0625	0.0000
चमक$_n^3$	0.0625	0.2543	0.1494	0.3231	0.0000
चमक$_n^4$	0.0625	0.3295	0.5097	0.0942	0.0000
कष्ट$_n^1$	0.0625	0.3295	0.5097	0.0942	0.0000
कष्ट$_n^1$	0.1250	1.0708	0.1195	0.1764	0.0083
झेलना	.0625	0.3238	0.1195	0.2398	0.0000

Table IV. Values of Local Connectivity Measures for the Sentence Graph Shown in Figure 6 (Using Fuzzy Hindi WordNet)

Words	Degree	PageRank	HITS	Closeness	Betweenness
चमक$_n^1$	0.0000	0.0083	0.0000	0.0555	0.0000
चमक$_n^2$	0.0000	0.0083	0.0000	0.0555	0.0000
चमक$_n^3$	0.1058	0.7098	0.2567	0.3284	0.4190
चमक$_n^4$	0.0588	0.5737	0.1199	0.2008	0.0000
कष्ट$_n^1$	0.1058	0.9611	0.1967	0.2700	0.2476
कष्ट$_n^2$	0.1176	0.8042	0.1679	0.2544	0.1333
झेलना	0.0588	0.5537	0.1199	0.2324	0.0000

some extra edges that represent the fuzzy entailment relation between चमक$_n^3$(camak^3_n) and मेहनत(mēhanata) and the fuzzy entailment relation between मेहनत(mēhanata) and कष्ट$_n^1$(kaṣṭa$_n^1$), i.e., चमकना(camakna) $\xrightarrow{0.8}$ मेहनत करना(mēhanata karanā), मेहनत करना(mēhanata karanā) $\xrightarrow{0.8}$ कष्ट सहना(kaṣṭa sahanā).

After constructing sentence graphs using both Hindi WordNet and Fuzzy Hindi WordNet, we compute the local connectivity measures for both the graphs using Equations (3) through (15). Table III shows the results of local connectivity measures for all vertices in the sentence graph (shown in Figure 5) constructed using Hindi WordNet. Table IV shows the results of local connectivity measures for all vertices in the sentence graph (shown in Figure 6) constructed using Fuzzy Hindi WordNet.

Now for each interpretation of the sentence, we construct the interpretation graph. The interpretation graph is the subset of the sentence graph. For each interpretation, we construct the interpretation graph from the sentence graph constructed using Hindi WordNet (Figure 5) and also construct the interpretation graph from the sentence graph constructed (Figure 6) using Fuzzy Hindi WordNet. The given sentence has three content words "चमक(camaka)," "कष्ट(kaṣṭa)," and "झेलना(jhēlanā)" for which Hindi WordNet lists 4, 2, and 1 senses, respectively.

Therefore, the given sentence has $I = \prod_{i\leq 1}^{n} m_i = 2 \times 1 \times 4 = 8$ interpretations. Thus, we construct $G_j(V_j, E_j) \subseteq G(V, E)(1 \leq j \leq 8)$, initializing $V_j, E_j(1 \leq j \leq 8)$:

V_1 = (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), V_2 = (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)),
V_3 = (चमक$_n^2$(camaka$_n^{2)}$, कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), V_4 = (चमक$_n^2$(camaka$_n^2$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)),
V_5 = (चमक$_n^3$(camaka$_n^3$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), V_6 = (चमक$_n^3$(camaka$_n^3$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)),
V_7 = (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), V_8 = (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)).
$E_j = \emptyset \ (1 \leq j \leq 8)$

We first construct the interpretation graphs using Hindi WordNet, i.e., from the sentence graph shown in Figure 5. First we construct the interpretation graph for the interpretation (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)). It may be noted that there is no path between node pairs चमक$_n^1$(camaka$_n^1$) and कष्ट$_n^1$(kaṣṭa$_n^1$), चमक$_n^1$(camaka$_n^1$) and झेलना(jhēlanā), and कष्ट$_n^1$(kaṣṭa$_n^1$) and झेलना(jhēlanā). Therefore, the constructed interpretation graph is disconnected for interpretation (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)). For a disconnected graph, the value of all global connectivity measures (compactness, graph entropy, edge density) are zero. Similarly, the interpretation graph constructed for the rest of the seven interpretations are also disconnected. Therefore, all global connectivity measures for all interpretation graphs are zero.

Now we construct the interpretation graph for each of these eight interpretations using Fuzzy Hindi WordNet, i.e., the subgraphs of the sentence graph shown in Figure 6. First we construct the interpretation graph for the interpretation (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)). There is no path between node pairs चमक$_n^1$(camaka$_n^1$) and कष्ट$_n^1$(kaṣṭa$_n^1$), चमक$_n^1$(camaka$_n^1$) and झेलना(jhēlanā), and कष्ट$_n^1$(kaṣṭa$_n^1$) and झेलना(jhēlanā). Therefore, the constructed interpretation graph is disconnected for interpretation (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)). Similarly, for interpretations (चमक$_n^1$(camaka$_n^1$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)), (चमक$_n^2$(camaka$_n^2$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), and (चमक$_n^2$(camaka$_n^2$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)), the interpretation graphs constructed are the disconnected, and thus a zero value of all global connectivity measures are obtained. For the rest of the four interpretations, namely (चमक$_n^3$(camaka$_n^3$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), (चमक$_n^3$(camaka$_n^3$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)), (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), and (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)), the interpretation graphs are constructed and shown in Figure 7(a) through (d), respectively. For the sake of brevity, we have shown the interpretation graphs only for the interpretations for which the graph constructed are not disconnected, and thus a nonzero value of global connectivity measures are obtained. Table V shows the resulting values of global connectivity measures for the interpretation graphs shown in Figure 7(a) through (d) for various interpretations of the sentence "*कष्ट झेलकर ही चमकते है* (*kaṣṭa jhēlkar hi camakatē hai*)" using Fuzzy Hindi WordNet.

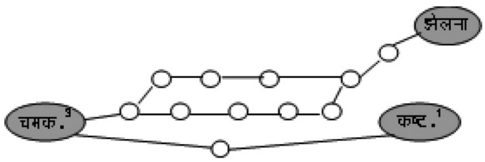


Fig. 7(a). Interpretation graph for interpretation (चमक$_n^3$, कष्ट$_n^1$, झेलना) from sentence graph in Figure 6.

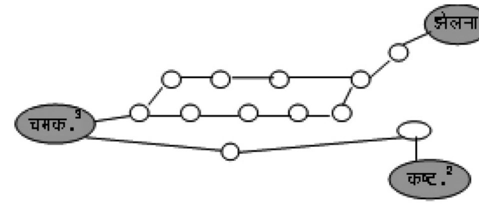


Fig. 7(b). Interpretation graph for interpretation (चमक$_n^3$, कष्ट$_n^2$, झेलना) from sentence graph in Figure 6.

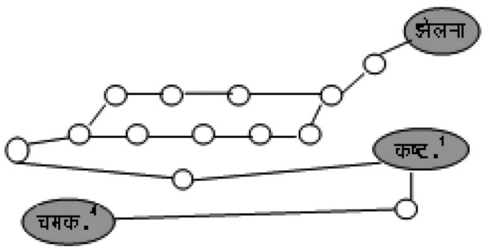


Fig. 7(c). Interpretation graph for interpretation (चमक$_n^4$, कष्ट$_n^1$, झेलना) from sentence graph in Figure 6.

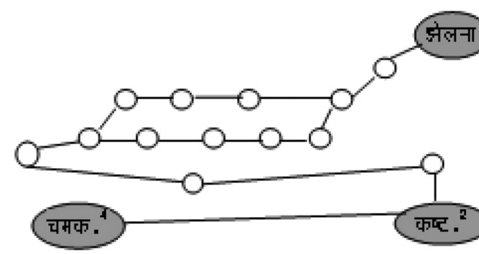


Fig. 7(d). Interpretation graph for interpretation (चमक$_n^4$, कष्ट$_n^2$, झेलना) from sentence graph in Figure 6.

Table V. Values of Global Connectivity Measures for the Sentence "*कष्ट झेलकर ही चमकते है*" using Fuzzy Hindi WordNet

Interpretation	Compactness	Graph Entropy	Edge Density
चमक$_n^3$ कष्ट$_n^1$ झेलना (Fig. 7(a))	0.9092	0.9817	0.1494
चमक$_n^3$ कष्ट$_n^2$ झेलना (Fig. 7(b))	0.9064	0.9334	0.1390
चमक$_n^4$ कष्ट$_n^1$ झेलना (Fig. 7(c))	0.9027	0.9429	0.1300
चमक$_n^4$ कष्ट$_n^2$ झेलना (Fig. 7(d))	0.9027	0.9429	0.1300

Now we compare the results for WSD of the sentence "*कष्ट झेलकर ही चमकते है (kaṣṭa jhēlkar hi camakatē hai)*" by using Hindi WordNet and Fuzzy Hindi WordNet.

First we discuss the results when Hindi WordNet is used. Table III shows the local connectivity measures, the degree centrality shows a tie between interpretations (चमक$_n^3$(camaka$_n^3$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)) and (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)), whereas PageRank centrality results in favor of the interpretation (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)). HITS favors interpretation (चमक$_n^4$(camaka$_n^4$), कष्ट$_n^1$(kaṣṭa$_n^1$), झेलना(jhēlanā)), whereas closeness centrality favors the interpretation (चमक$_n^3$(camaka$_n^3$), कष्ट$_n^2$(kaṣṭa$_n^2$), झेलना(jhēlanā)). Betweenness centrality, the last local measure, is unable to find any conclusion. Being zero values of all global connectivity measures (using Hindi WordNet), it is not possible to give any results. Now it may be easily noted that maximum support goes in favor of the interpretation (चमक$_n^4$(camaka$_n^4$) कष्ट$_n^2$(kaṣṭa$_n^2$) झेलना(jhēlanā)) using Hindi WordNet. Therefore, the final interpretation using Hindi WordNet is "*दुख झेलकर ही टीस होती है। (dukh jhēlkar hi ṭīsa hoti hai)*," which is a wrong interpretation.

Now we discuss the results when Fuzzy Hindi WordNet is used. All local connectivity measures, namely PageRank, HITS, closeness, and betweenness centrality (except degree centrality) support interpretation (चमक$_n^3$(camaka$_n^3$) कष्ट$_n^2$(kaṣṭa$_n^1$) झेलना(jhēlanā)) (see Table IV). Degree centrality supports चमक$_n^3$(camaka$_n^3$) for content word "चमक(camaka)," which is correct, whereas कष्ट$_n^2$(kaṣṭa$_n^1$) for the content word "कष्ट(kaṣṭa)" with a small difference of 0.01, which may be avoided. The global connectivity measures (compactness, entropy, edge density) also favor the interpretation (चमक$_n^3$(camaka$_n^3$) कष्ट$_n^2$(kaṣṭa$_n^1$) झेलना(jhēlanā)) (see Table V). Therefore, both the local and the global connectivity measures results in a (चमक$_n^3$(camaka$_n^3$) कष्ट$_n^2$(kaṣṭa$_n^1$) झेलना(jhēlanā)) interpretation. Taking this interpretation, the sentence "*कष्ट झेलकर ही चमकते है।* (*kaṣṭa jhēlakara hī camakte hai*)" can be written as "*यंत्रणा झेलकर ही विकास होता है।* (*yantranā jhēlkar hi vikāsa hota hai*)," which is the correct interpretation. Therefore, by employing Fuzzy Hindi WordNet, we are able to find the correct interpretation of the Hindi sentence, whereas Hindi WordNet failed to identify the correct interpretation of the Hindi sentence.

6. TIME COMPLEXITY, EXPERIMENTAL SETUP, AND RESULTS

In this section, we first compute the time complexity of our algorithm. We also compare the performance of the proposed method by using Hindi WordNet and Fuzzy Hindi WordNet. We find that Fuzzy Hindi WordNet gives better results than Hindi WordNet.

6.1. Time Complexity

We discuss the time complexity of the proposed WSD method by using Hindi WordNet and Fuzzy Hindi WordNet. Our WSD method consists of two steps: the first is graph construction and the second is disambiguation.

Let s be the constant that represents the highest number of senses of any word w in Hindi WordNet (HWN)/Fuzzy Hindi WordNet ($FHWN$).

$$s = max_{w \in HWN/FHWN} |Senses\,(w)\,| \tag{24}$$

Let the sentence σ have n words. In the associated sentence graph, the number of senses is bound by $|V\sigma|$, where

$$|V\sigma| = |\cup_{i=1}^{n} Senses\,(w_i)\,| \leq \sum_{i=1}^{n} |Senses\,(w_i)| \leq \sum_{i=1}^{n} s = sn \tag{25}$$

For the sentence σ, the graph construction procedure executes DFS from each vertex $v \in Vs$. the running time for DFS from each vertex is $O(|V_{FHWN}| + |E_{FHWN}|)$, where V_{FHWN}, E_{FHWN} are the sets of vertices and edges, respectively, in the entire Fuzzy Hindi WordNet graph. The number of edges $|E_{FHWN}|$ can be bound by $k \times |V_{FHWN}|$, where k is a constant equal to the maximum number of Fuzzy Hindi WordNet edges incident to any node in the Fuzzy Hindi WordNet graph. So the cost of a single DFS is $O(|V_{FHWN}| + k|V_{FHWN}|) = O(|V_{FHWN}|)$. Given that the number of vertices in $Vs$is bound by $sn \in O(n)$, the overall time complexity of the graph construction phase is $O(n|V_{\mathrm{FHWN}}|)$. In practice, however, the running time is nearly equal to $O(n^2)$. Since we do not explore the entire Fuzzy Hindi WordNet graph, a DFS can take $O(n)$ time. The maximum number of incident edges is a small value, and we visit the Fuzzy Hindi WordNet graph at the maximum distance $\leq L$.

With a local measure lm, the running time of the disambiguation phase is $O(C_{lm}(n))$, where $C_{lm}(n)$ is the time cost incurred by lm when applied to all $O(n)$ nodes in the sentence graph. Table VI shows the running times for each measure individually. It can be seen that the degree complexity amounts to $O(n)$ time if we represent the sentence graph by an adjacency list. Measures of eigenvector centrality require the application

Table VI. Time Complexity of Local and Global Measures

	Local Measure					Global Measure		
	Degree	PageRank	HITS	KPP	Betweenness	Compactness	Graph Entropy	Edge Density
Runtime	$O(n)$	$O(n^2)$	$O(n^2)$	$O(n^2)$	$O(nm)$	$O(\mu n^2)$	$O(\mu n)$	$O(\mu n)$

of the power method, which takes $O(n^2)$ time. The time complexity of the key player problem [Johnson 1997] depends on the calculation of all shortest distances if m is the number of edges in the sentence graph and $m \in O(n)$; it costs $O(m + n)$ for each vertex, as each vertex has a constant upper bound on the number of incident edges. Finally, for the betweenness local measure, researchers explain an $O(nm)$ implementation [Brandes 2001].

Complexity increases when global measures are employed. Calculating the score of a single graph (corresponding to one interpretation of sentence) takes $O(n^2)$ time ($O(n)$ for edge density and graph entropy if we use an adjacency list). Exhaustively generating all possible interpretations is computationally prohibitive with $O(s^n)$ complexity (recall that s is the maximum number of senses for a word in Fuzzy Hindi WordNet). We can reduce the search space to a very large constant μ using the approximation algorithms. The running time of the approximated global measures is thus polynomial.

It may be noted that representation of the Hindi WordNet and Fuzzy Hindi WordNet graphs can only be differentiated by the number of edges. For a given set of words, more words/nodes are related in Fuzzy Hindi WordNet than in Hindi WordNet. In addition to all connections/edges defined in Hindi WordNet, Fuzzy Hindi WordNet has some more additional edges representing fuzzy relations. Thus, the number of edges increase in the Fuzzy Hindi WordNet graph by a constant value as compared to the Hindi WordNet graph. The number of edges $|E_{FHWN}|$ is bound by $k \times |V_{FHWN}|$, where k is a constant equal to the maximum number of Fuzzy Hindi WordNet edges incident to any node in the Fuzzy Hindi WordNet graph. Thus, the time complexity of the proposed WSD method increases only by a constant when Fuzzy Hindi WordNet is used in place of Hindi WordNet, and hence the order of time complexity remains the same.

6.2. Experimental Setup and Results

We evaluated our algorithm using the publicly available sense marked corpus on the IIT Bombay Web site (Center for English Language Technology 2010). Some of the sentences in the health corpus and various fuzzy relations existing between the words of the sentences are shown in Table VII. For any given system, the scorer reports precision (the number of correct senses over the number of senses returned), recall (the number of correct senses over the total number of senses identified in the evaluation dataset), and their combined F_1 measure = (2PR/(P+R)). Since our method provides an answer for all ambiguous words, precision, recall, and F_1 are the same. We report the performance for Fuzzy Hindi WordNet and Hindi WordNet. Since our method provides an answer for all ambiguous words, precision, recall, and F_1 are the same. Our results on the health corpus subset are summarized in Table VIII. We report the performance for Hindi WordNet and Fuzzy Hindi WordNet. The All column shows results on all words (monosemous and polysemous), whereas the Poly column shows results on the polysemous words subset. We compare the results by using both Hindi WordNet and Fuzzy Hindi WordNet reference lexicons one by one.

Let us first focus on the results we obtained with the standard Hindi WordNet computational lexicon. From Table VIII, it can easily be seen that all local measures perform better than the global measures. Out of the local measures, degree and PageRank give better results than the other three. However, degree and PageRank give almost similar results (as can be seen in Table VIII). This is consistent with the previous results obtained by researchers using these two measures [Upstill et al. 2003]. Among the global

Table VII. Sentences Extracted from Sensed Marked Hindi Corpus and Existing Fuzzy Relations between the Words

	Hindi Sentence	Pre-processed Open class words	Existing Fuzzy Relation between open class words
1.	इसमें मुख के प्रदाह सहित छोटे पीड़ादायक छाले होते हैं	मुख प्रदाह छोटा पीड़ादायक छाला	मुख Fuzzy Meronyny छाला छाला Fuzzy Attribute पीड़ादायक
2.	इसमें सलादों के रूप में एक दो ताजी कच्ची सब्जियाँ और कम से कम दो पकी हुई सब्जियाँ हो सकती हैं	सलाद ताजी कच्चा सब्जी कम पका	सब्जी Fuzzy Attribute कच्चा सब्जी Fuzzy Attribute पका सलाद Fuzzy meronymy सब्जी
3.	में गन्दा स्वाद, आलेपित जीभ तथा बदबूदार श्वास	अन्य लक्षण मुख गन्दा स्वाद आलेपित जीभ बदबूदार श्वास	स्वाद Fuzzy Attribute गन्दा जीभ Fuzzy Attribute आलेपित श्वास Fuzzy Attribute बदबूदार
4.	कभी कभी गले में घुटन की संवेदना का अनुभव किया जाता है	गला घुटन संवेदना अनुभव	घुटन Fuzzy Hypernymy अनुभव
5.	अनेक लोग जो अपना खाना तनाव में तथा शीघ्रता से निगलते हैं इस बीमारी से पीडित होते हैं	अनेक लोग खाना तनाव शीघ्रता निगलना बीमारी	तनाव Fuzzy Meronymy बीमारी
6.	यह त्वचा को कोमल बने रहने में सहायता करेगा	त्वचा कोमल सहायता	त्वचा Fuzzy Attribute कोमल
7.	आँतों को क्रियाशील बनाने के लिए टहलना या धीरे धीरे दौड़ना प्रारंभ दौड़ना प्रारंभ किया जाना चाहिए	आँत क्रियाशील टहलना धीरे धीरे दौड़ना प्रारंभ	टहलना Fuzzy Antonymy दौड़ना
8.	मिरगी एक बहुत प्राचीन रोग है जिसने नेपोलियन, एलेक्जेंडर तथा जूलियस सीजर सहित विश्व के बहुत से महानतम व्यक्तियों को पीडित किया	मिरगी बहुत प्राचीन रोग सहित विश्व बहुत व्यक्ति महानतम पीडित	व्यक्ति Fuzzy Attribute महानतम
9.	कच्ची हरी सब्जियाँ इस खनिज से समृद्ध होती हैं	कच्ची हरी सब्जी समृद्ध होना	सब्जी Fuzzy Attribute कच्ची सब्जी Fuzzy Attribute हरी
10.	कच्ची सब्जियों के रस, विशेषकर अलग से या चुकन्दर और खीरे के रस के साथ मिलाकर लिया गया गाजर का रस, थकान से उबरने में अत्यधिक मूल्यवान हैं	कच्ची सब्जी रस विशेषकर अलग चुकन्दर खीरा मिलाना गाजर रस थकान अत्यधिक मूल्यवान	सब्जी Fuzzy Attribute कच्ची चुकन्दर Fuzzy Meronymy रस खीरा Fuzzy Meronymy रस गाजर Fuzzy Meronymy रस
11.	मसाले और बघार, तथा अचार, जो भोजन को और अधिक स्वादिष्ट बनाते हैं तथा अधिक खाने के लिए लेना बाध्य करते हैं, से अवश्य दूर रहना चाहिए	मसाला बघार अचार भोजन अधिक स्वादिष्ट अधिक खाना बाध्य अवश्य दूर	अचार Fuzzy Meronymy मसाला बघार Fuzzy Meronymy मसाला भोजन Fuzzy Meronymy मसाला भोजन Fuzzy Meronymy अचार भोजन Fuzzy Meronymy बघार
12.	सलाद ड्रेसिंग में नीबू के रस का प्रयोग किया जा सकता है	सलाद नीबू रस प्रयोग	सलाद Fuzzy Meronymy नीबू
13.	निद्रा शरीर के लिए विश्राम की एक नियतकालिक अवस्था है	निद्रा शरीर विश्राम अवस्था	विश्राम Fuzzy Meronymy निद्रा
14.	इस आहार नियम में, उसे केले को छोड़कर सेब, नाशपाती, अंगूर, संतरे तथा अनन्नास जैसे ताजे रसदार फलों का दिन में तीन बार भोजन करना चाहिए	टाहार नियम केला सेब नाशपाती अंगूर संतरे अनन्नास ताजे रसदार फल दिन भोजन	आहार Fuzzy Association भोजन भोजन Fuzzy Meronymy फल
15.	रोगी को ताजी सब्जियों तथा फलों के रसों का प्रचूर मात्रा में सेवन करना चाहिए	रोगी ताजी सब्जी रस प्रचूर मात्रा सेवन	सब्जी Fuzzy Attribute ताजी
16.	उसकी बदबूदार साँस, पीलियाग्रस्त त्वचा और उदर में फूली हुई शिराएँ होती है	बदबूदार साँस पीलियाग्रस्त त्वचा उदर फूली शिराए	साँस Fuzzy Attribute बदबूदार
17.	जब आप बीमार या अस्वस्थ अनुभव करें तो चिंता न करें	बिमार अस्वस्थ अनुभव चिंता	बीमार Fuzzy Association अस्वस्थ
18.	विश्राम तथा निद्रा थकान पर काबू पाने और ऊर्जा की पुनः प्राप्ति में सहायता करते हैं	विश्राम निद्रा थकान ऊर्जा	विश्राम Fuzzy Meronymy निद्रा
19.	दैनिक स्नान शरीर को मैल तथा गंध से मुक्त रखता है	दैनिक स्नान शरीर मैल गंध	मैल Fuzzy Hypernymy गंध
20.	इसे आमंत्रित न करें, सोने की उत्कट इच्छा न करें, बस लेट जाएँ और आराम करें	आमंत्रित सोना उत्कट इच्छा लेटना आराम	सोना Fuzzy Entails लेटना लेटना Fuzzy Entails आराम करना

Table VIII. Performance of Connectivity Measures (F_1 Measure)

		Hindi WordNet		Fuzzy Hindi WordNet	
		All	Poly	All	Poly
Local	Degree	0.5307	0.4102	0.6183	0.4931
	PageRank	0.5290	0.4067	0.6072	0.4905
	HITS	0.4508	0.3240	0.5281	0.4376
	KPP	0.4801	0.3654	0.5598	0.4453
	Betweeness	0.4890	0.3708	0.5923	0.4850
Global	Compactness	0.4232	0.2953	0.4863	0.3962
	Graph entropy	0.4127	0.2803	0.4562	0.3120
	Edge density	0.4257	0.3012	0.5192	0.4153

measures, compactness and edge density are better than graph entropy. It can easily be seen that local measures provide better results than global measures and that all differences are statistically significant.

We now discuss the performance of the different measures for WSD using Fuzzy Hindi WordNet. It can be seen that the best local measures are degree, PageRank, and betweenness, as their values are almost similar. The best global measure is edge density, which outperforms compactness and graph entropy. However, all local measures are significantly better than edge density. A reference lexicon with a large number of semantic relations, i.e., a greater number of edges seems to benefit both local and global measures. Fuzzy Hindi WordNet yields better results compared to Hindi

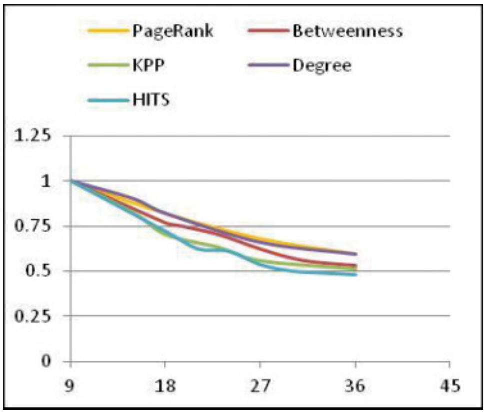


Fig. 8(a). Performance of local connectivity measures for average-length sentences.

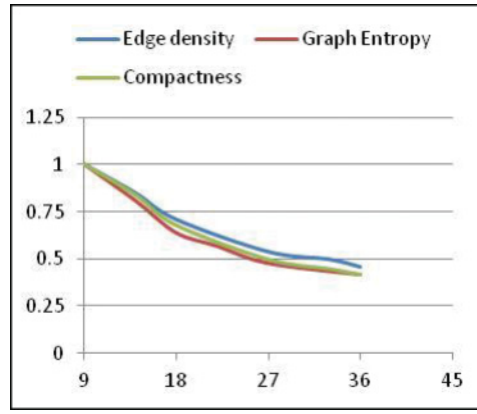


Fig. 8(b). Performance of global connectivity measures for average-length sentences

WordNet. By using Fuzzy Hindi WordNet sense inventory, the disambiguation performance increases in most cases by approximately 8%.

The health corpus contains 90 files, and each file contains approximately 100 sentences. The average number of open-class words in the sentences in health corpus is 8.85 (we take the nearest integer, i.e., 9). The average number of word senses for the words defined in Hindi WordNet is 3.63 (we take the nearest integer, i.e., 4). As the average word senses in Hindi WordNet is 4, the total number of word senses in the average length sentence may vary from 9 to $9 \times 4 = 36$. We extracted all sentences with an average word length from the health corpus and evaluated the performance of the proposed method for those sentences. Figure 8 shows the performance of the proposed algorithm in terms of precision. The outcomes of the experimentations given in Figure 8 show that the proposed method is scalable for sentences having an average word length.

Overall, we find that WSD can be achieved at a higher level by using Fuzzy Hindi WordNet in place of Hindi WordNet.

7. CONCLUSIONS AND FUTURE WORK

Fuzzy Hindi WordNet represents the relations between words as a matter of degree, as most pairs of words in the real world may be partially related. We found that the relations in Hindi WordNet convey more natural representation when extended to Fuzzy Hindi WordNet. The representation of Fuzzy Hindi WordNet using fuzzy graphs provides a generalization of the relations by assigning membership degree. Further, the relations between the words that are not directly related can be inferred using composition of fuzzy relations. The strength of composed relations can easily be computed by selecting the appropriate t-norm. We proposed and used fuzzy graph connectivity measures (both local and global) for WSD. We found that Fuzzy Hindi WordNet gives better results for WSD than that shown using Hindi WordNet. Fuzzy Hindi WordNet can provide improved results in information retrieval and query expansion, as query can be effectively processed using fuzzy relations such as fuzzy association, fuzzy hypernymy/hyponymy, fuzzy meronymy/holonymy, and fuzzy entailment. Since the fuzzy relations defined in Fuzzy Hindi WordNet are helpful in bringing closer more semantically related words and receding away less semantically related words, it may be useful in better understanding of context while using machine translation. Since reference lexicons like WordNet exist for several languages (EuroWordNet, MultiWordNet, etc.), an interesting future direction would be to establish and evaluate the performance of extending WordNet to Fuzzy WordNet with respect to these resources. In the future, the performance of Sentiwordnet [Esuli and Sebastiani 2006] (a lexical resource based on WordNet explicitly devised for supporting sentiment classification and opinion mining) can be improved by including fuzzy relations in it. All words in Sentiwordnet are not assigned a polarity value, and the fuzzy relations and composition of fuzzy relations proposed in this article may be helpful in automatically assigning polarities to each word, which is otherwise a tedious task that requires intensive manual effort. It may be helpful in areas such as social network tracking, health, psychology, and music. Since WordNet has been used for several different purposes in information systems, including ducument classification, automatic text summarization, and to determine the similarity between words, this newly introduced idea of Fuzzy WordNet and the composition of fuzzy relations can be used in these applications in the future.

GLOSSARY

Hindi Word	Corresponding English Meaning	Hindi Word	Corresponding English Meaning	Hindi Word	Corresponding English Meaning
आम(āma)	Mango	पर्स(parsa)	Purse	तैरना(tairanā)	Swim
गुठली(guṭhalī)	Kernel	बैग(baiga)	Bag	मछली(machalī)	Fish
लडकी(laḍakī)	Girl	छात्रवास(chātravāsa)	Boys Hostel	पेंगुइन(pēnguina)	Penguin
औरत(aurata)	Woman	गत्ता(gattā)	A pasteboard	मगरमच्छ(magaramaccha)	Crocodile
विद्यालय(vidyālaya)	School	किताब(kitāba)	Book	वाहन चलाना(vāhana calānā)	To drive vehicle
संस्थान(saṁsthāna)	Institute	रेडियो(rēiḍayō)	Radio	पढाना(paṛhānā)	To teach
पेड (pēḍa)	Tree	कार(kāra)	Car	उपकुलपति(upakulapati)	Vice chancellor
फल(phala)	Fruit	कॉलर(kêālara)	Collar	प्रोफेसर(pṛōphēsara)	Professor
अलमारी(alamārī)	Almirah	कमीज(kamīza)	Shirt	सहप्रोफेसर(sahapṛōphēsara)	Associate Prof.
रैक(raika)	Rack	ऑले(ôlē)	Ice balls in rain	सहायक प्रोफेसर(sāhāyāk pṛōphēsara)	Asst. Prof.
वाहन(vāhana)	Vehicle	बारिश(bāriśa)	Rain	शोध करना(śōdha karanā)	To research
कार(kāra)	Car	ज्वारभाटा(jvārabhāṭā)	Jwar Bhata	प्रशासन(pṛśāsana)	Administration
पहिया(pahiyā)	Wheel	समुन्द्र(samundṛ)	Ocean	सुंदर(sundara)	Beautiful
बेईमान(bēīmāna), बदनीयत(badanīyata) ईमानफरोश (īmānapharōśa)	Dishonest	खुश(khuśa)	Haappy	जमीन(jamīna)	Land
सज्जन(sajjana), भला आदमी(bhalā ādamī), शरीफ(śarīpha), सुजन(sujana), साहु(sāhu)	Gentle	ठंडा(ṭhaṇḍā)	Cold	उपजाऊ(upajāū)	Fertile
लडका(laḍakā)	Boy	गुनगना(gunagunā)	Luke warm	चॉंदी(côdī)	Silver
आदमी(ādmī)	Man	गर्म(garma)	Hot	दिन(dina)	Day
चूडी(cūḍī)	Thin bangle of glass	हॅसना(hêsanā)	To Laugh	साफ(sāpha)	Clean
कंगन(kangana)	Bangle	कराहना(karāhanā)	Groan	पानी(pānī)	Water
मेज(mēja)	Table	बडा(badā)	Big	गर्म (garma)	Hot
स्टूल(sṭūla)	Stool	थोडा(thōḍā)	Few	सुपात्र(supātra)	Worthy
किताब(kitāba)	Book	ज्यादा(jyādā)	More	व्यक्ति(vyakti)	Person
कुंजी(kuñjī)	Guide Book	छोटा(chōṭā)	Small	तेज(tēja)	Fast
दस्ता(dastā)	A quire of paper	धुंधला(dhundhalā)	Blurred	भागना(bhāganā)	Run
कॉपी (kêāpī)	Copy	चमकदार(camakadāra)	Shiny	क्रमशः(kṛmaśaḥ)	Serially
जीप(jīpa)	Jeep	सफलता मिलना (saphalatā milanā)	To get success	क्रम(kṛma)	Serial
स्कूटर(skūṭara)	Scooter	मेहनत करना(mēhanata karanā)	To work hard	चलाते(calātē)	To drive
मोटरसाईकिल(mōṭarasāīkila)	Motorbike	सोना(sōnā)	Sleep	समय(samaya)	Time
उदास(udāsa)	Sad	लेटना(lēṭanā)	Lie down	सुनने(sunnē)	To listen
चिंतित(cintita)	Tensed	बदबू आना(badabū ānā)	To feel foul Smell	आनन्द(ānnada)	Delight
अटैची(aṭaicī)	Briefcase	गंदगी होना(gandagī hōnā)	To be dirty	अलग(alaga)	Different
संदूक(sandūka)	Trunk	अच्छा व्यवहार होना (acchā vyavahāra hōnā)	To behave well	चमक(camaka)	Shininess/ Sparkle/ To get success/ Pang
दुखी(dukhī)	Miserable	अच्छे संसकार होना(acchē saṁskāra hōnā)	To have good moral	कष्ट	Anguish/ Sadness
परेशान(parēśāna)	Annoyed	स्वस्थ होना(svastha hōnā)	To be healthy	झेलना/झेलकर (jhēlanā/jhēlakara)	To bear

दयालु(dayālu)	Kind	व्यायाम करना(vyāyāma karanā)	To exercise	दुख(dukh)	Grief
ईमानदार(īmānadāra)	Honest	चमकना(camakanā)	To shine	टीस(ṭīsa)	Pain
छाछ(chācha)	Buttermilk	कष्ट सहना(kaṣṭa sahanā)	To bear pain	यंत्रणा(yantranā)	Anguish
दही(dahī)	Curd	दुखी होना(dukhī hōnā)	To be sad	विकास(vikāsa)	Success
अध्यापक(adhyāpaka)	Teacher	परेशान होना(parēśāna hōnā)	To be tensed		
आईसक्रीम(āīskṛīma)	Ice-cream	दौडना(dauḍanā)	To run		
दूध(dūdha)	Milk	चलना(calanā)	To walk		
आवाज(āvāja)	Voice	पढना(paḍhanā)	To read		
ध्वनि(dhvani)	Sound	सीखना(sīkhanā)	To Learn		
पेपर.कटर(pēpara kaṭara)	Paper Cutter	हॅसना(hêsanā)	To Laugh		
हथियार(hathiyāra)	Weapon	सामान्य मनोदशा (sāmānya manōdaśā)	Normal mental state		
क्रूर(kṛūra)	Cruel	कराहना(karāhanā)	To moan		
दुर्जन(durjana)	Rogue	खाना(khānā)	To eat		
दुष्ट(duṣṭa)	Wicked	खिलाना(khilānā)	To serve food		

REFERENCES

N. Alon and A. Orlitsky. 1996. Source coding and graph entropies. *IEEE Transactions on Information Theory* 42, 5, 1329–1339.

K. Avneet. 2010. Development of an approach for disambiguating ambiguous Hindi post position. *International Journal on Computer Applications* 5, 9, 25–32.

S. Banerjee and T. Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using WordNet. In *Computational Linguistics and Intelligent Text Processing*. Lecture Notes in Computer Science, Vol. 2276. Springer, 136–145.

J. C. Bezdek, G. Biswas, and L. Huang. 1986. Transitive closures of fuzzy thesauri for information-retrieval systems. *International Journal of Man-Machine Studies* 25, 343–356.

P. Bhattacharya. 1987. Some remarks on fuzzy graphs. *Pattern Recognition Letters* 6, 5, 297–302.

P. Bhattacharyya. 2010. IndoWordNet. In *Proceedings of the Lexical Resources Engineering Conference*.

P. P. Bonissone and S. D. Keith. 1986. Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity. *Machine Intelligence Pattern Recognition* 4, 17–247.

S. P. Borgatti. 2006. Identifying set of key players in a social network. In *Computational and Mathematical Organization Theory*. Springer, 21–34.

R. A. Botafogo, E. Rivlin, and B. Shneiderman. 1992. Structural analysis of hypertexts: Identifying hierarchies and useful metrics. *ACM Transactions on Information Systems* 10, 2, 142–180.

U. Brandes. 2001. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology* 25, 2, 163–177.

Center for Indian Language Technology. 2010. Hindi Corpus. Retrieved October 20, 2015, from http://www.cfilt.iitb.ac.in/Downloads.html.

S. M. Chen, Y. J. Horng, and C. H. Lee. 2001. Document retrieval using fuzzy valued concept network. *IEEE Transactions on Systems, Man and Cybernetics* 31, 1, 111–118.

S. M. Chen, Y. J. Horng, and C. H. Lee. 2003. Fuzzy information retrieval based on multi relationship fuzzy concept network. *Fuzzy Sets and Systems* 140, 1, 183–205.

M. D. Cock, S. Guadarrama, and M. Nikravesh. 2005. Fuzzy thesauri for and from the WWW. In *Soft Computing for Information Processing and Analysis*. Studies in Fuzziness and Soft Computing, Vol. 164. Springer, 1–10.

S. Das, A. Seetha, M. Kumar, and J. L. Rana. 2010. Post translation query expansion using Hindi WordNet for English-Hindi CLIR system. In *Proceedings of the Forum for Information Retrieval Evaluation (FIRE'10)*.

C. Diou, G. Katsikatsos, and A. Delopoulos. 2006. Constructing fuzzy relations from WordNet for word sense disambiguation. In *Proceedings of the 1st Workshop on Semantic Media Adaptation and Personalization*. 135–140.

S. K. Dwivedi and P. Rastogi. 2008. An entropy based method for removing Web query ambiguity in Hindi language. *Journal of Computer Science* 4, 9, 762–767.

K. Erk, D. McCarthy, and N. Gaylord. 2013. Measuring word meaning in context. *Computational Linguistics* 39, 3, 511–554.

A. Esuli and S. Fabrizio. 2006. SENTIWORDNET: A publically available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation* (*LREC'06*). 417–422.

C. Fellbaum. 1998. *WordNet*: *An Electronic Lexical Database*. MIT Press, Cambridge, MA.

L. C. Freeman. 1979. Centrality in social networks conceptual clarification. *Social Networks* 1, 3, 215–239.

L. C. Freeman, S. P. Boragatti, and D. R. White. 1991. Centrality in valued graph: A measure of betweenness based on network flow. *Social Networks* 13, 141–154.

S. Fujita and A. Fujino. 2013. Word sense disambiguation by combining labeled data expansion and semisupervised learning method. *ACM Transactions on Asian Language Information Processing* 12, 2, Article No. 7.

G. K. Gupta. 2006. *Introduction to Data Mining with Case Studies*. Prentice Hall of India Pvt. Ltd., 238–240.

A. Jain, K. Mittal, and D. K. Tayal. 2014a. Automatically incorporating context meaning for query expansion using graph connectivity measures. *Progress in Artificial Intelligence* 2, 2–3, 129–139.

A. Jain, D. K. Tayal, and S. Rai. 2014b. Shrinking digital gap through automatic generation of WordNet for Indian languages. *Open Forum* 30, 2, 215–222. DOI:10.1007/s00146-014-0548-5

A. Jain, D. K. Tayal, and S. Yadav. 2014c. Retrieving Web search results for a Hindi query using max-max soft clustering. *International Journal of System Assurance Engineering and Management*, 1–12.

A. Jain, S. Yadav, and D. Tayal. 2013. Measuring context meaning for open class words in Hindi language. In *Proceedings of the 6th International Conference on Contemporary Computing* (*IC3*). 118–123.

D. B. Johnson. 1977. Efficient algorithms for shortest paths in sparse networks. *Journal of the ACM* 24, 1–13.

J. M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46, 604–632.

G. J. Klir and T. A. Folger. 1988. *Fuzzy Sets, Uncertainty and Information*. Prentice Hall.

G. K. Kotle and S. C. Bhairal. 2009a. Exploring links in WordNet hierarchy for word sense disambiguation of nouns. In *Proceedings of the International Conference on Advances in Computing, Communication, and Control.* 20–25.

G. K. Kotle and S. C. Bhairal. 2009b. WordNet: A knowledge source for word sense disambiguation. *International Journal of Recent Trends in Engineering* 2, 4, 213–217.

R. Kumar and G. N. Jha. 2010. Translating politeness across cultures: Case of Hindi and English. In *Proceedings of the 3rd International Conference on Intercultural Collaboration*. ACM, New York, NY, 175–178.

S. Kumar and V. Mansotra. 2012. Query optimization: A solution for low recall problem in Hindi language information retrieval. *International Journal of Computer Applications* 55, 17, 6–17.

K. H. Lee. 2005. *First Course on Fuzzy Theory and Applications*. Springer-Verlag, Berlin.

B. Luca and C. Carlos. 2006. The distribution of pagerank follows a power-law only for particular values of the damping factor. In *Proceedings of the 15th International Conference on World Wide Web*. ACM, New York, NY, 941–942.

K. Martin. 1992. A fuzzy concept network model and its application. In *Proceedings of 1st IEEE International Conference on Fuzzy Systems*.

S. Mathew and M. S. Sunitha. 2009. Types of arcs in a fuzzy graph. *Information Sciences* 179, 1760–1768.

R. Mihalcea. 2005. Unsupervised large vocabulary word sense disambiguation with graph based algorithm for sequence data labelling. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing* (*HLT/EMNLP'09*). 411–418.

R. Mihalcea and R. Dragmoir. 2011. *Graph-Based Natural Language Processing*. Cambridge University Press.

G. A. Miller and C. Fellbaum. 2007. WordNet then and now. *Language Resources and Evaluation* 41, 2, 209–214.

N. Mishra, S. Yadav, and T. J. Siddiqui. 2009. An unsupervised approach to Hindi word sense disambiguation. In *Proceedings of the 1st International Conference on Intelligent Human Computer Interaction* (*IHCI'09*). 327–335.

D. Narayan, D. Chakrabarty, P. Pande, and P. Bhattacharyya. 2002. An experience in building the Indo WordNet: A WordNet for Hindi. In *Proceedings of the 1st International Conference on Global WordNet*.

R. Navigli. 2005. Semiautomatic extension of large scale linguistic knowledge bases. In *Proceedings of the 18th Florida Artificial Intelligence Research Society Conference*. 548–553.

R. Navigli. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys* 41, 2, 1–69.

R. Navigli and M. Lapata. 2010. An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 4, 678–692.

M. E. J. Newman. 2005. A measure of betweenness centrality based on random walks. arXiv:cond-mat/0309045.

T. Opsahl, F. Agneessens, and J. Skvoretz. 2010. Node centrality in weighted network generalising degree and shortest path. *Social Networks* 32, 3, 245–251.

S. P. Ponzetto and R. Navigli. 2010. Knowledge-rich word sense disambiguation rivalling supervised systems. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. 1522–1531.

F. P. Romero, P. Julian-Iranzo, A. Soto, M. Ferreira-Satler, and J. Gallardo-Casero. 2012. Classifying unlabeled short texts using a fuzzy declarative approach. *Language Resources and Evaluation* 47, 1, 151–178.

A. Rosenfeld, L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura. 1975. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*. Academic, New York, NY, 77–97.

K. V. Sandeep and K. V. Chanchal. 2012. A graph based word sense disambiguation for Hindi language. *International Journal of Scientific Research Engineering and Technology* 1, 5, 313–318.

A. Sharan, M. L. Joshi, and A. Pandey. 2011. Exploiting ontology for concept based information retrieval. In *Information Systems for Indian Languages*. Springer, Berlin, 157–164.

V. Sheinman, C. Fellbaum, I. Julien, P. Schulam, and T. Tokunaga. 2013. Erratum to: Large, huge or gigantic? Identifying and encoding intensity relations among adjectives in WordNet. *Language Resources and Evaluation* 47, 3, 817–818.

M. Sinha, M. K. Reddy, R. P. Bhattacharya, P. Pandey, and L. Kashyap. 2004. Hindi word sense disambiguation. In *Proceedings of the International Symposium on Machine Translation, Natural Language Processing, and Translation Support Systems*.

S. Sinha and T. J. Siddiqui. 2012. Evaluating effect of context window size, stemming and stop word removal on Hindi word sense disambiguation. In *Proceedings of the Information Retrieval and Knowledge Management International Conference* (*CAMP'12*). IEEE, Los Alamitos, CA, 1–5.

M. S. Sunitha. 2001. *Studies on Fuzzy Graph*. Ph.D. Dissertation. Cochin University of Science and Technology, Cochin, India.

D. K. Tayal, P. C. Saxena, A. Sharma, G. Khanna, and S. Guptz. 2014. New method for solving reviewer assignment problem using type-2 fuzzy sets and fuzzy functions. *Applied Intelligence* 40, 1, 54–73.

T. Upstill, N. Craswell, and D. Hawking. 2003. Predicting fame and fortune: PageRank or indegree? In *Proceedings of the 8th Australasian Document Computing Symposium*. 31–40.

D. Widdow and B. Dorow. 2002. A graph model for unsupervised lexical acquisition. In *Proceedings of the 19th International Conference on Computational Linguistics*. 1–7.

R. R. Yager. 2010. Concept representation and database structure in fuzzy social relational networks. *IEEE Transactions on Systems, Man and Cybernetics—Part A*: *Systems and Humans* 40, 2, 413–419.

J. Yen and R. Langari. 2005. *Fuzzy Logic Intelligence, Control and Information*. Pearson Education.

L. A. Zadeh. 1965. Fuzzy sets. *Information and Control* 8, 3, 338–353.

L. A. Zedeh, K. S. Fu, and M. Shimura. 1975. Fuzzy relations, fuzzy graphs, and their applications to clustering analysis. In *Fuzzy Sets and Their Application to Cognitive and Decision Processes*. Academic Press, New York, NY, 125.

Received December 2013; revised November 2014; accepted June 2015