

Instructions for Authors of WDCOPIN

Rodrigo Souza¹

¹Departamento de Sistemas e Computação – Universidade Federal de Campina Grande (UFCG)
Caixa Postal 10.106 – CEP 58.109-970 – Campina Grande – PB – Brazil

rodrigo@dsc.ufcg.edu.br

Abstract.

Resumo.

1. Introdução

A divisão conceitual de um sistema de software em módulos é uma informação valiosa durante o seu desenvolvimento, sobretudo para a divisão de tarefas entre desenvolvedores [cite]. Essa informação muitas vezes não é documentada adequadamente e, à medida que os desenvolvedores são substituídos, pode se perder.

Algoritmos de modularização de software procuram identificar a organização de um sistema de software a partir de sua implementação e, por isso, auxiliam desenvolvedores a recuperar a organização modular de um sistema sobre o qual se tem pouco conhecimento. Para realizar a divisão em módulos, alguns algoritmos usam heurísticas baseadas em identificadores, ou então informações sobre o histórico de desenvolvimento (CVS). Muitos algoritmos, no entanto, consideram apenas o conjunto de dependências entre os componentes da implementação do sistema.

Sistemas de software, no nível de implementação, são formados por vários componentes que interagem entre si. Programas orientados a objetos, por exemplo, são compostos por classes que interagem com outras classes através de mecanismos como herança e troca de mensagens. As interações estabelecem relações de dependência entre os componentes, de forma que cada componente só funciona corretamente na presença dos componentes dos quais ele depende.

A rede de dependências entre componentes de um sistema pode ser extraída automaticamente através de ferramentas de análise estática de código-fonte ou código objeto. A rede extraída deve ser considerada apenas uma aproximação da rede de dependências real, pois nenhuma técnica automática é capaz de determinar corretamente, em todos os casos, se existe uma dependência entre dois componentes [cite].

Estudos recentes analisaram redes de dependências entre componentes sob a luz da teoria das redes complexas e encontraram propriedades topológicas marcantes que também estão presentes em redes de diversos domínios, como redes de interações entre proteínas e redes sociais.

2. Identificação do Problema

Avaliação é importante. Dizer como é feita a avaliação de algoritmos de recuperação de arquitetura. Dificuldade em se avaliar algoritmos de recuperação (por causa da escassez de sistemas cuja estrutura modular é documentada).

3. Objetivos

Resumir em um parágrafo a nossa abordagem de redes sintéticas.

O OBJETIVO DE PESQUISA É**ANALISAR MODELOS ESTOCÁSTICOS QUE POSSAM SER USADOS PARA AVALIAR ALGORITMOS DE RECUPERAÇÃO DE ARQUITETURA - Descobrir modelos estocásticos de redes livres de escala organizadas em módulos. - Caracterizar redes de software. - Desenvolver método para atestar a "softwaressimilhança" de uma rede. Avaliar esse método. Ele deve atestar que redes de software são softwaressímeis, enquanto redes de interações entre proteínas não são. - Avaliar os modelos quantitativamente quanto à verossimilhança das redes de software que eles geram. - Aplicar algoritmos de recuperação de arquitetura a redes sintéticas, calcular authoritativeness e comparar resultados com aqueles disponíveis na literatura. Estudo exploratório.

Descobrir modelos estocásticos capazes de sintetizar sistemas de software organizados em módulos e avaliá-los quantitativamente quanto à capacidade de gerar sistemas semelhantes a sistemas reais (sobretudo desenvolver o método para avaliação).

(E, principalmente: Desenvolver um método!) Objetivo: demonstrar que a hipótese X é verdadeira

4. Atividades Planejadas e Realizadas

Realizadas: - Produção de uma revisão bibliográfica - Produção de um manuscrito com resultados preliminares da pesquisa - Implementação de modelos - Coleta de um corpo de sistemas de software - Coleta de um corpo de redes não-software - Desenvolvimento de um método de avaliação de realismo (e aplicação a um corpus pequeno) – DESCREVER O MÉTODO sucintamente. - Avaliação qualitativa e quantitativa de modelos (2 modelos não-orientados) através de um experimento.

Planejadas: - Refinamento do método de avaliação de realismo. Est.: - Avaliar modelos de redes orientadas. - Achar uma métrica que não dependa de achar softwares de mesmo tamanho. (e que se aplique a redes orientadas, uma vez que em redes orientadas a distância de Andrade gera matrizes muito esparsas, afetando negativamente a relevância do valor calculado) - Publicação com resultados preliminares. Est.: outubro? - Aplicação de algoritmos de clustering e comparação com literatura: Est.: - Dissertação - Publicação dos resultados finais da pesquisa

5. Resultados Obtidos

Desenvolvemos o modelo BCR+, que produz redes orientadas modulares que satisfazem a uma arquitetura modular fornecida como parâmetro. Ele foi baseado em um modelo de redes livres de escala orientadas sem módulos [cite]. Encontramos ainda dois outros modelos de redes orientadas modulares, o LR e o CGW, e um modelo de rede modular não-orientada, o LFR.

Desenvolvemos um procedimento experimental para responder à seguinte pergunta: os modelos são capazes de produzir redes semelhantes a redes extraídas de sistemas reais? Consideramos que um modelo é bem sucedido se ele é capaz de sintetizar uma rede cuja distância para uma rede real tomada como referência não seja maior do que a distância entre redes reais com o mesmo número de vértices da rede de referência.

Nesse experimento tomamos como referência a rede do jogo VilloNanny 2.2.4, com 1658 vértices. Avaliamos os modelos BCR+ e LR.

avaliar o realismo das redes de software sintéticas, baseado na distância de Andrade. Esse procedimento foi executado para avaliar o realismo. O procedimento foi descrito em um artigo ainda não publicado.

Os dados mostram que o modelo BCR+ foi o que mais se aproximou da rede de referência; uma das redes geradas ficou mais próxima à rede de referência do que a rede do sistema JFreeChart. O modelo LFR produziu redes mais próximas entre si, como revela o pequeno desvio-padrão, porém mais distantes da rede de referência.

Estudo qualitativo e quantitativo sobre a verossimilhança dos modelos. Mostrar resultados experimentais: - Medição de frequência de motifs. - Mostrar tabela de distâncias. - Mostrar tabela das métricas (distribuição de graus etc). - Mostrar gráfico da distribuição de motifs e mostrar que há uma diferença visual.